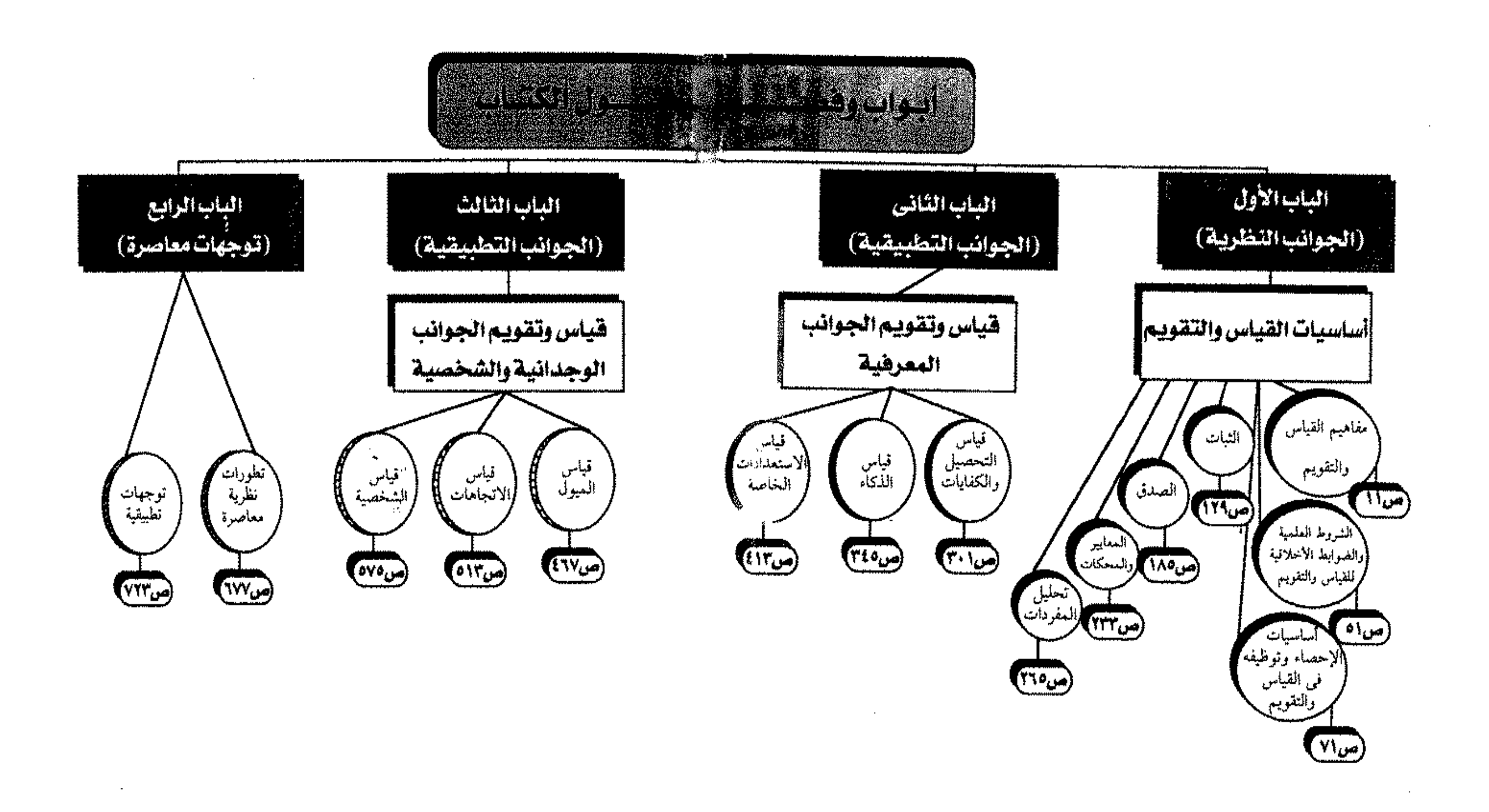
القياس والتقويم التربوي والنفسي

أساسياته وتطبيقاته وتوجهاته المعاصرة

الدكتور صلاح الدين محمود علام





الباب الأول
(الجوانب النظرية)
أساسيات القياس والتقويم
مفاهيم القياس والتقويم
ص١١
الشروط العلمية والضوابط الأخلاقية للقياس والتقويم
ص٥١
أساسيات الإحصاء وتوظيفه فى القياس والتقويم
ص٧١
الثبات
ص١٢٩
الصدق
ص١٨٥
المعايير والمحكات
ص٢٣٣
تحليل المفردات
ص٢٦٥
الباب الثانى
(الجوانب التطبيقية)
قياس وتقويم الجوانب المعرفية
قياس التحصيل والكفايات
ص٣٠١
قياس الذكاء
ص٣٤٥
قياس الاستعدادات الخاصة
ص٤١٣
الباب الثالث
(الجوانب التطبيقية)
قياس وتقويم الجوانب الوجدانية والشخصية
قياس الميول
ص٤٦٧
قياس الاتجاهات
ص٥١٣
قياس الشخصية
ص٥٧٥
الباب الرابع
(توجهات معاصرة)
تطورات نظرية معاصرة
ص٦٧٧
توجهات تطبيقية
ص٧٢٣

القياس والتقويم التربوي والنفسي

أساسياته وتطبيقاته وتوجهاته المعاصرة

الدكتور صلاح الدين محمود علام
أستاذ القياس والتقويم والإحصاء التربوي والنفسي
كلية التربية - جامعة الأزهر

الطبعة الأولى
١٤٢٠ هـ - ٢٠٠٠ م

ملتزم الطبع والنشر
دار الفكر العربي
٩٤ شارع عباس العقاد - مدينة نصر - القاهرة
ت: ٢٧٥٢٩٨٤ - فاكس: ٢٧٥٢٧٣٥

٣٧١,٢٦ صلاح الدين محمود علام.

ص ل ق ى القياس والتقويم التربوى والنفسى: أساسياته وتطبيقاته وتوجيهاته المعاصرة/ تأليف صلاح الدين محمود علام. - القاهرة: دار الفكر العربى، ١٩٩٩.

٨١٦ص: إيض؛ ٢٤سم.

ببليوجرافية: ص ٧٥٣-٧٧٦.

يشتمل على معجم بالمصطلحات إنجليزى - عربى.

تدمك: ٠ - ١٢١٠ - ١٠ - ٩٧٧.

١ - الإختبارات والقياسات التعليمية. أ- العنوان.

تصميم وإخراج فنى

السيد أحمد سيف

أميرة للطباعة عابدين - ت : ٣٩١٥٨١٧

بسم الله الرحمن الرحيم

مقدمة

يُعد مجال القياس والتقويم التربوى والنفسى من المجالات الحيوية الأساسية التى لا غنى عنها للدارسين والباحثين فى العلوم السلوكية، والمسئولين عن اتخاذ القرارات المتعلقة بالأفراد فى مختلف الميادين التطبيقية التربوية والنفسية والاجتماعية والإدارية والصناعية، والعسكرية، وغيرها من الميادين التى تتمركز حول الإنسان مصدر الثروة الرئيسة للأمم الواعية.

لذلك حظى مجال القياس والتقويم التربوى والنفسى باهتمام بحثى تطويرى متزايد من جانب علماء النفس والتربية فى الآونة الأخيرة. ولعل هذا يبدو واضحًا وملموسًا فيما تتناوله أدبيات القياس المعاصر المتخصصة من موضوعات بحثية مستحدثة، وما ترفُده من إسهامات تطبيقية فى مجالات السلوك الإنسانى المختلفة.

ولم يعُد يُنكر أحد اليوم الدور البالغ الأهمية للقياس التربوى والنفسى فى تقدم العلوم السلوكية وإثراء تطبيقاتها الوظيفية المتعددة بما يفى بالمتطلبات المتجددة للحياة الإنسانية فى هذا القرن.

غير أن تنوع مجالات القياس والتقويم التربوى والنفسى وما يندرج تحتها من موضوعات متعددة، وما تنطوى عليه من مفاهيم وأساليب وأدوات متباينة، يُؤدى فى كثير من الأحيان إلى إساءة استخدام هذه الأساليب والأدوات وعدم توظيفها توظيفًا مناسبًا. وربما يرجع ذلك إلى عدم وضوح أوجه الترابط والتكامل فيما بينها نظرًا لحاجة الدارسين والباحثين والممارسين إلى تطوير معارفهم فيما يتعلق بالأسس المنهجية والإحصائية التى تستند إليها، وتنمية مهاراتهم فى تطبيق مفاهيم وأساليب وأدوات القياس والتقويم المعاصر بفهم مستنير وفكر متفتح.

لذلك بدت الحاجة ماسة إلى مرجع أساسى متكامل يكون بمثابة مرشد ينير الطريق للدارسين والباحثين والممارسين فى مختلف مجالات القياس والتقويم التربوى والنفسى.

وقد حاولنا في هـذا الكتاب أن يكون هناك توازن وتكامل بين الجوانب النظرية والجوانب التطبيقية بحيث يخدم كلاهما الآخر مُراعين الأسس المنهجية والإحصائية التي ترتكن إليها أساليب وأدوات القياس والتقويم، والاستخدامات المناسبة لها، وحدود هذه الاستخدامات، لكي تفي بالغرض الذي صُممت من أجل تحقيقه. واسترشدنا في عرضنا لهذه الجوانب بأهم التطورات التي حدثت في علم القياس والتقويم التربوي والنفسي لمواكبة الحركة البحثية المعاصرة، كما أفدنا من خبرتنا الأكاديمية وممارساتنا الميدانية في مجال القياس والتقويم والإحصاء في كثير من الدول العربية خلال العقدين الماضيين. وتيسيراً على القارئ قسمنا الكتاب إلى أربعة أبواب رئيسة متكاملة، **تناولنا في الباب الأول** الجوانب النظرية المتعلقة بالمفاهيم والمبادئ الأساسية، والأسس المنطقية والإحصائية للقياس والتقويم التربوي والنفسي.

وتناولنا في البابين الثاني والثالث أهم التطبيقات المتعلقة بقياس وتقويم الجوانب المعرفية والوجدانية والشخصية. **وتناولنا في الباب الرابع** بعض التطورات والتوجهات المعاصرة في هذا المجال.

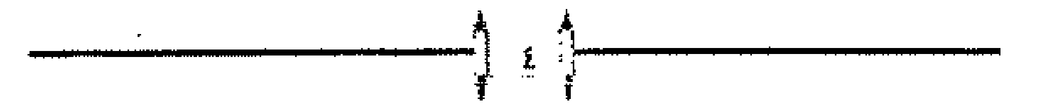
ويشتمل الباب الأول على سبعة فصول، يتناول الفصل الأول المفاهيم والمبادئ الأساسية للقياس والتقويم، وأنواع السمات النفسية والتربوية وخصائصها، واستخدامات أدوات القياس والتقويم، وتصنيفها.

ويتناول الفصل الثاني مشكلات القياس التربوي والنفسي وشروطه العلمية، وضوابط وأخلاقيات استخدام الاختبارات والمقاييس.

ويتناول الفصل الثالث أساسيات الإحصاء وتوظيفه في القياس والتقويم التربوي والنفسي متضمنًا أساليب تنظيم وعرض البيانات وتمثيلها بيانيًا أو بشكل الأغصان والأوراق، ومقاييس النزعة المركزية والتشتت، ووصف توزيعات درجات الاختبارات، والأنواع المختلفة من الدرجات المحوّلة، ومقاييس العلاقة بين مجموعتين من الدرجات، واستخدام معادلة الانحدار الخطي البسيط في التنبؤ.

ويتناول الفصل الرابع الخصائص السيكومترية للاختبارات المتعلقة بثبات الدرجات، ومصادر أخطاء القياس، والأنواع المختلفة لمعاملات الثبات، وكيفية تقدير قيمها، وثبات تقديرات المحكمين، وثبات الفروق بين درجتي اختبارين، والخطأ المعياري للقياس، والعوامل المؤثرة في قيم معامل الثبات.

ويتناول الفصل الخامس الخصائص السيكومترية للاختبارات المتعلقة بالصدق وجوانبه، وطرق تقدير صدق المحتوى، والصدق المرتبط بمحك بنوعيه التنبؤي



والتلازمى، والعوامل التى تؤثر فيه، وكذلك صدق التكوين الفرضى ومنهجيته، وأساليب جمع أدلة متعلقة به، ومنظور تكاملى لطرق تقدير الصدق والثبات.

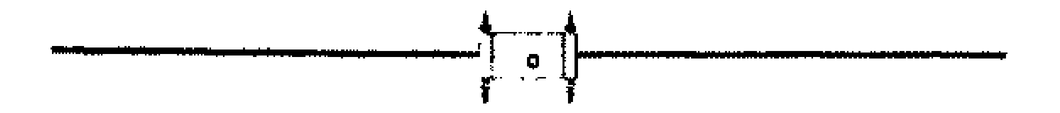
ويتناول **الفصل السادس** معايير الاختبارات مرجعية الجماعة أو المعيار، ومحكات الأداء فى الاختبارات مرجعية المحك، وأنواع الجماعات المرجعية وخصائصها، وأنواع معايير الاختبارات، والصفحات النفسية والتربوية «البروفيل»، واستخدام محكات الأداء فى تفسير الدرجات.

ويتناول **الفصل السابع** أساليب تحليل مفردات الاختبارات وأهمية هذا التحليل، وكيفية تقدير صعوبة المفردات والمشكلات المتعلقة بذلك، وأساليب تقييم تمييز المفردات استنادًا إلى محكات داخلية أو خارجية، وتحليل فاعلية بدائل الإجابات، واستخدام تحليل المفردات فى انتقاء مفردات الاختبار، واستخدم الحاسوب فى إجراء هذا التحليل.

ويشتمل **الباب الثانى** - المتعلق بقياس وتقويم الجوانب المعرفية - على ثلاثة فصول، يتناول **الفصل الثامن** قياس التحصيل والكفايات، وعلاقة مفهوم التحصيل بمفهوم الاستعداد، والتمييز بين اختبارات كل منهما، وتصنيف اختبارات التحصيل واستخداماتها، وخطوات بناء اختبارات التحصيل مرجعية المحك، واستخدامات هذه الاختبارات، واختبارات التحصيل المقننة (جماعية الجماعة أو المعيار)، وبعض أوجه النقد التى توجّه لاختبارات التحصيل المقننة.

ويتناول **الفصل التاسع** قياس الذكاء العام، ومفاهيم الذكاء المتعددة والمشكلات المتعلقة بذلك، واختبارات الذكاء العام الفردية والجماعية، وأمثلة متنوعة لكل من نوعى الاختبارات، وأساليب بناء اختبارات الذكاء، وبعض القضايا المتعلقة بقياس الذكاء،وبعض محددات الذكاء والعوامل المؤثرة فيه، وثبوت نسبة الذكاء.

ويتناول **الفصل العاشر** قياس الاستعدادات الخاصة، وعلاقة مفهوم الاستعداد بمفهوم القدرة، وتأثير العوامل الوراثية والبيئية فى الاستعدادات الخاصة، واستخدامات اختبارات الاستعدادات، وبعض الافتراضات الأساسية فى قياسها، وأساليب بناء اختباراتها، وإجراءات تحليل العمل، وبطاريات الاستعدادات المتعددة، واختبارات الاستعدادات المتعلقة بمجالات نوعية محددة، مثل الاستعداد الميكانيكى والاستعدادات النفسية الحركية، والاستعداد الكتابى، والاستعداد الموسيقى، والاستعدادات الفنية، والابتكارية، ومفهوم التفكير التباعدى، وقياس قدرات هذا النمط من التفكير، والمشكلات المتعلقة به.

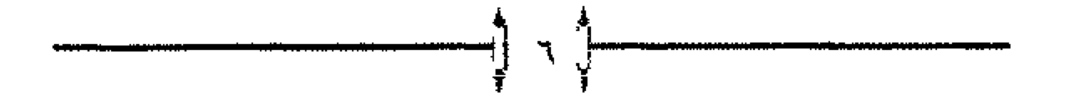


ويشتمل الباب الثالث ـ المتعلق بقياس وتقويم الجوانب الوجدانية والشخصية ـ على ثلاثة فصول، يتناول الفصل الحادى عشر قياس الميول، وعلاقة مفهوم الميل بمفهوم الاستعداد، والأنماط المختلفة للميول، وأساليب بناء استبيانات الميول، وبعض المشكلات المتعلقة بذلك، وأمثلة لهذه الاستبيانات واستخداماتها، وتقويم ميول الطلبة في الصف الدراسي، وأدوات قياس هذه الميول.

ويتناول الفصل الثاني عشر قياس الاتجاهات، وعلاقة مفهوم الاتجاه بغيره من المفاهيم الوجدانية، ومكونات الاتجاهات وخصائصها، والأسس المنطقية لقياسها، وتصنيف مقاييس الاتجاهات، وأساليب بناء موازين الاتجاهات، مثل أسلوب ثيرستون، وأسلوب ليكرت، وأسلوب جتمان، وأسلوب تمايز معاني المفاهيم، وأسلوب التصنيف الترتيبي (Q Sort)،وأسلوب البسط، وأساليب قياس العلاقات الاجتماعية، وأساليب بناء موازين الاتجاهات متعددة الأبعاد، وإرشادات عامة لبناء موازين الاتجاهات، وأنواع أخرى من مقاييس الاتجاهات، وبعض مشكلات قياس الاتجاهات واستخدامات مقاييسها.

ويتناول الفصل الثالث عشر قياس الشخصية، ومفهوم الشخصية وبعض نظرياتها، وتصنيف مقاييسها، وطرق وأساليب قياسها، واستراتيجيات بناء استبيانات الشخصية، وأمثلة لبعض أهم هذه الاستبيانات التي تستند إلى أساس نظري منطقي، أو إلى التحليل العاملي، أو إلى أساس إمبيريقي، وكذلك استبيانات جبرية الاختيار، وتقييم استبيانات الشخصية بعامة، وتصنيف الأساليب الإسقاطية، وأمثلة لبعض أهم اختباراتها وتقييمها، وأساليب ملاحظة السلوك، ومصادر الخطأ فيها وكيفية التغلب عليها، وموازين تقدير السلوك، ومصادر الخطأ فيها ومعالجتها، وأساليب المقابلة وميزاتها وعيوبها وتقييمها، ومقاييس الأنشطة الفسيولوجية، ومقاييس الإدراك وتقييمها، وتعقيب عام على طرق وأساليب قياس الشخصية، واستخدامات هذه المقاييس.

ويشتمل الباب الرابع على فصلين، يتناول الفصل الرابع عشر بعض التطورات النظرية المعاصرة، ومن أهمها نظرية الاستجابة للمفردة الاختبارية، والأنواع الرئيسة لنماذجها، وأساسها المنطقي، وتصنيفها،ومفهوم المنحنى المميز للمفردة، والنماذج الاستاتيكية اللوغاريتمية أحادية البعد، مثل نموذج راش، ونموذج البارامترين، ونموذج البارامترات الثلاثة، ومزايا هذه النماذج وعيوبها، ونظرية الثبات متعدد الأوجه، وبعض مفاهيمها الأساسية، والدراسات المتعلقة بها، وتحليل مكونات التباين في هذه الدراسات، والعلاقة بين هذه النظرية ومفهوم الصدق، وكذلك يتناول إسهام علم النفس



المعرفى فى تطوير أساليب القياس والتقويم، وعلاقة النظرية المعرفية بالعملية الاختبارية، وتصميم الاختبارات فى ضوء هذه النظرية.

ويتناول الفصل الخامس عشر بعض التوجهات التطبيقية المتعلقة بأساليب قياس وتقويم التحصيل الدراسى، واستخدام الاختبارات مرجعية المحك فى تقويم التحصيل، وتقنيات تفريد قياس التحصيل باستخدام الاختبارات المفصلة، وبناء نظم بنوك الأسئلة، ومراحل إنشاء نظام بنك أسئلة مقترح، والبنك الدولى للأسئلة، وتقويم التحصيل التربوى من منظور دولى، والتقييم التوثيقى الشامل للتحصيل والكفايات Portfolio Assessment ، ومكونات الصحائف الوثائقية Portfolios وصيغها ، وبعض المشكلات المتعلقة بها واقتراحات للتغلب عليها.

وقد قدمنا فى بداية كل فصل من فصول البابين الثانى والثالث عرضًا موجزاً لتطور قياس السمة التى يتناولها الفصل، لكى يتضح للقارئ هذا التطور من منظور تاريخى.

كما قدمنا العديد من الأشكال التخطيطية التى تيسر استيعاب المفاهيم والأساليب المتعلقة بقياس مختلف السمات الإنسانية، وكذلك قدمنا فى بداية الكتاب شكلاً تخطيطيًا يرشد القارئ إلى أبواب الكتاب وفصوله والصفحات المناظرة لها.

وينتهى الكتاب بمجموعة منتقاة من المراجع التى يمكن الرجوع إليها لمزيد من التفاصيل فى موضوعات الكتاب المتعددة.

ويسعدنى أن أقدم جزيل الشكر للأستاذ المهندس / **عاطف محمود الخضرى** مدير دار الفكر العربى على اهتمامه وحفزه للانتهاء من هذا الكتاب وفق جدول زمنى محدد يدل على سداد تفكيره وفاعلية إدارته.

ونرجو من الله سبحانه وتعالى أن يفيد بهذا الكتاب الدارسين والباحثين والممارسين فى مختلف مجالات القياس والتقويم التربوى والنفسى، وأن يمكنهم من اكتساب المهارات الأساسية المتعلقة بذلك، وتطوير هذه المهارات بحيث تستوعب التطورات المعاصرة المتسارعة التى حدثت فى هذا العلم الإنسانى الحيوى.

والله نسأل أن يحقق الرجاء

صلاح الدين محمود علام

دكتوراه الفلسفة Ph.D.

فى التقويم والقياس والإحصاء النفسى والتربوى

من جامعة ميتشجان الأمريكية



أساسيات القياس والتقويم التربوى والنفسى



الفصل الأول

المفاهيم والمبادئ الأساسية

* مقدمة
* أهمية القياس العلمى
* المفهوم العلمى للقياس
* القياس التربوى والنفسى
* طبيعة القياس التربوى والنفسى
* أنواع القياس ومستوياته
* مدخل السمات فى القياس التربوى والنفسى
* بعض الخصائص الأساسية للسمات
* أنواع السمات النفسية والتربوية
* عمليات القياس وأدوات القياس
* الاختبار التربوى والنفسى
* التقويم والتقييم
* استخدامات أدوات القياس والتقويم
* تصنيف الاختبارات والمقاييس

مقدمة :

يهتم علماء النفس والتربية بدراسة السلوك الإنسانى دراسة علمية من أجل فهم أفضل لخصائص هذا السلوك وتوجيهه لخير الإنسان، ورفاهية مجتمعه. ونحن عندما نتعامل مع الآخرين، نود أيضًا معرفة تفضيلاتهم ومشاعرهم وإمكاناتهم. ورغبتنا فى معرفة الآخرين تكون أحيانًا لمجرد حب الاستطلاع، وأحيانًا أخرى لتعرُّف كيفية التعامل معهم فى سياق الحياة اليومية. وتعتمد هذه المعرفة فى غالبية الأحيان على الملاحظة الذاتية والآراء الشخصية، مما يجعلها تتسم بعدم الاتساق أو الموضوعية وربما التحيز. ولكن المعلم فى أدائه لمهامه المهنية، وفى تعامله مع تلاميذه،يهتم بسلوكهم وميولهم وتوجهاتهم وإنجازاتهم وما يمكنهم إنجازه مستقبلاً لكى تحقق العملية التعليمية أهدافها المرجوة. وكذلك القائمين بدراسة إمكانات القوى البشرية فى مختلف مواقع العمل، يهتمون بتعرف خصائص الأفراد من أجل الإفادة القصوى من إمكاناتهم ، ورفع مستوى أدائهم وإنتاجيتهم تلافيًا للهدر فى الجهد والوقت والمال. وفى مثل هذه الحالات ينبغى اتخاذ قرارات لا يشوبها الحدس، أو الذاتية،أو التحيز، وإنما يجب أن تستند إلى الأساليب العلمية فى الانتقاء والتصنيف، والفحص والتشخيص ،والتسكين والترفيع. ومن هنا تبرز أهمية اكتساب هؤلاء المعنيين المعارف الأساسية، والمهارات الإجرائية المتعلقة بالقياس والتقويم التربوى والنفسى.

وتحقيقًا لذلك سوف نفرد هذا الباب بفصوله السبعة لمناقشة الجوانب النظرية متضمنة الأسس المنطقية والإحصائية والمنهجية للقياس والتقويم بما يزود القارئ بذخيرة من المعلومات تستند إليها التطبيقات التى سوف نقدمها فى البابين الثانى والثالث، وبذلك يستطيع الإفادة منها فى تنمية مهاراته المرجوة.

ولعل من المناسب أن نستهل هذا الفصل بعرض أهم المفاهيم والمصطلحات والحقائق الأساسية المتعلقة بمجال القياس والتقويم التربوى والنفسى.

أهمية القياس العلمي:

يعد القياس أمرًا على جانب كبير من الأهمية في أي علم من العلوم، فجميع العلوم تسعى لتطوير أساليب موضوعية دقيقة لقياس الظواهر المتعلقة بها من أجل فهم هذه الظواهر وتفسيرها، والتنبؤ بالعلاقات القائمة بين متغيراتها، ومحاولة ضبطها والتحكم فيها. فالتقدم العلمي يعتمد إلى حد كبير على تمثيل الظواهر والأحداث وصياغتها بأساليب موضوعية دقيقة، بحيث تمكن الدارسين والباحثين والممارسين من التواصل فيما بينهم بلغة مشتركة متفق عليها، كما تمكنهم من التقييم الموضوعي للنتائج التي يتوصل إليها العلماء، وبدون هذه الأساليب يصبح تعريف المفاهيم والمصطلحات والقواعد المتعلقة بالظواهر المختلفة خاضعًا للآراء الذاتية ، والأحكام الفردية، وبذلك لا تكون هناك أسس متفق عليها في دراسة وبحث هذه الظواهر.

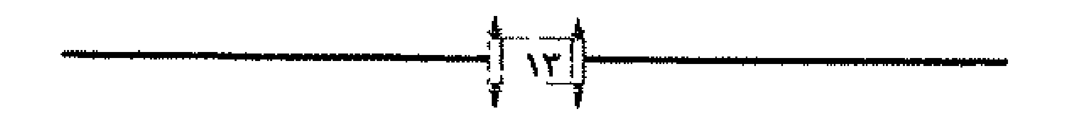
فالعلم نظام مفتوح نتائجه متداولة مما يستدعي أن تكون الإجراءات والاستنتاجات العلمية خالية من الغموض، بحيث يستطيع علماء مختلفون تكرارها، والتحقق من صدقها، وهذا لا يتأتى إلا إذا استندت هذه الإجراءات والاستنتاجات إلى أساليب قياس موضوعية.

وربما كان افتقار بعض العلوم في الماضي إلى مثل هذه الأساليب الموضوعية في القياس هو الذي أدى إلى التباين الواضح والجدل المستمر حول تعريف المفاهيم والمصطلحات والقواعد المتعلقة بالظواهر التي تهتم هذه العلوم بدراستها. ولعل علم النفس يعد مثالا واضحًا لذلك،فدقة وموضوعية القياس تسهم إذن في تحديد المفاهيم، وبلورة التفكير من أجل الفهم المستنير لطبيعة الظواهر المختلفة، وبدون ذلك تصبح الدراسة العلمية لهذه الظواهر خاضعة للتأملات العقلية والخبرات الشخصية.

ب- المفهوم العلمي للقياس:

إن القياس Measurement بمفهومه الواسع يشير إلى الجوانب الكمية التي تصف خاصة أو سمة معينة،مثل ارتفاع سائل ، أو حجم كرة، أو ضغط غاز، أو الاستعداد اللفظي لطفل، أو التحصيل الدراسي لطالب، كما يشير إلى عملية جمع المعلومات، وترتيبها بطريقة منظمة. وبذلك يتضمن مفهوم القياس من هذا المنظور كلا من عملية جمع المعلومات وتنظيمها، وكذلك نتيجة هذه العملية.

غير أنه يمكن تعريف القياس بصورة أكثر دقة بأنه: تعيين فئة من الأرقام أو الرموز، مناظرة لفئة من الخصائص أو الأحداث، طبقًا لقواعد محددة تحديدًا جيدًا.



ويشتمل هذا التعريف على ثلاثة عناصر هى : فئات الأرقام أو الرموز، وفئات الخصائص أو الأحداث، وقواعد تعيين الأرقام أو الرموز للخصائص أو الأحداث.

وقد استخدمنا كلمة فئة « الأرقام أو الرموز Numerals » بدلا من كلمة «أعداد Numbers» ، لأن الأعداد تشير إلى معانى الأرقام أو الرموز. فالعدد ٣٥ مثلاً يمكن أن يشير إلى رقم لاعب كرة، أو رقم تليفون، أو ترتيب طالب فى صف دراسى معين، أو درجة حرارة سائل، أو طول فرد. وهذه المعانى تتضمن علاقات مختلفة بين الأرقام التى تعد بمثابة رموز، والقياس يختص بهذه العلاقات.

أما فئة الخصائص أو الأحداث فتتعلق بمحتوى المجال العلمى، فعلم الكيمياء يتناول الغازات والسوائل، والأجسام الجاسئة من حيث الجزيئات والذرات والإلكترونات والشحنات، وعلم النفس يتناول السلوك الإنسانى من حيث التعلم والتحصيل، والميول والاستعدادات ، وهكذا فى بقية العلوم.

واختلاف فئات الخصائص أو الأحداث يقابله بالضرورة اختلاف فى نوع القياس. وهذا يتطلب تحديد قواعد تعيين الأرقام أو الرموز التى تناظر الخصائص أو الأحداث، بحيث يمكننا التعامل مع الأعداد بطرق معينة للتوصل من ذلك إلى استنتاجات نعيد تطبيقها على الظاهرة المقاسة.

فعندما نقول : إن طول محمد ١٧٠سم، فإننا نعلم تمامًا القواعد التى على أساسها أجرى قياس هذا الطول، فنحن نعلم أن هناك أداة لقياس الطول ذات تداريج موحدة ومتساوية تبدأ من الصفر؛ لذلك فإن العدد ١٧٥سم الذى يناظر خاصة أو صفة الطول أصبح له مدلول معين متعارف عليه ، ولا يختلف فى تفسيره أحد. فالصفر يعنى انعدام الطول، وهناك خصائص معينة للأعداد المدونة على المسطرة أو شريط القياس لها ما يناظرها فى ظاهرة الطول المراد قياسه، فكل من هذه الأعداد يتميز بالتفرد ؛ لأن الرقم (٥) يختلف عن الرقم (٢) مثلا، والطول المناظر لكل من هذه الأرقام ، يتميز أيضًا بالتفرد، فالشىء الذى طوله ٥ سم يتميز بأنه أطول فعلا من الشىء الذى طوله ٢سم.

والأعداد المدونة على الشريط تتميز بالترتيب ، وبذلك تؤدى إلى ترتيب الأطوال المناظرة لها، كما أن هذه الأعداد قابلة للإضافة، أى إذا جمعنا عددين أو أكثر ينتج عددًا آخر متميزًا،وكذلك الأطوال المناظرة لأى أعداد يمكن جمعها لتصبح طولا مساويًا لمجموع الأطوال المقاسة.

وبهذا الشكل تصبح الأعداد المدونة على الشريط المدرج الذى نقيس به ظاهرة الطول تمثل الأطوال تمثيلا رياضيًا له ما يناظره فى الواقع الفعلى. وبذلك يمكننا القول: إن هناك تماثلا أو تناظرًا أحادى بين خصائص ظاهرة الطول والنظام العددى المستخدم على الشريط المدرج الذى يفترض أنه يقيس الطول.

القياس التربوى والنفسى :

إن تعريف المفهوم العلمى للقياس الذى أشرنا إليه ينطبق أيضًا على القياس التربوى والنفسى ، إذ يمكن إعادة صياغة هذا التعريف لكى يناسب محتوى المجال التربوى والنفسى بحيث يصبح كالتالى:

القياس : هو تعيين فئة من الأرقام أو الرموز تناظر خصائص أو سمات الأفراد طبقًا لقواعد محددة تحديدًا جيدًا، وهذا يعنى أن القياس التربوى والنفسى يُعنى بتكميم خصائص أو سمات الأفراد، حيث إننا لا نستطيع قياس الأفراد فى ذاتهم، وإنما نقيس خصائصهم أو سماتهم.

فلكى نقيس خاصة أو سمة معينة لدى فرد وفقا لهذا التعريف، فإننا نحتاج إلى تحديد الفرد المراد قياس سمة معينة لديه، وتحديد السمة المراد قياسها، والإجراءات التى سوف تتبع لتعيين أرقام أو رموز Numerals تناظر السمة المقاسة. والأرقام الناتجة التى تكون عادة قيمًا عددية أو درجات Scores ينبغى - كما ذكرنا - أن تحافظ على العلاقات الفعلية القائمة بين مستويات السمة المقاسة لدى الأفراد.

وهذا يعنى أنه إذا كان فرد ما يتميز بقدرة لفظية أعلى من فرد آخر، فإن درجة الفرد الأول فى اختبار يقيس هذه القدرة ينبغى أن تكون أعلى من درجة الفرد الثانى فى الاختبار نفسه، أى أن الدرجات أو القيم العددية تقع على متصل تقاس عليه السمة. والهدف من عملية القياس تحديد موقع الفرد والسمة معًا على هذا المتصل، وهذا لابد أن يعتمد على إجراءات كافية ودقيقة بحيث يمكن اعتبار الدرجة التى يحصل عليها فرد فى اختبار، وهى مجموع النقاط التى ينالها فى أسئلة الاختبار، تكميم أو قياس لقدرته الفعلية مما يتطلب تقييمًا تجريبيًا . وسوف نتناول ذلك باستفاضة بعد قليل ، وكذلك فى الفصول اللاحقة.

طبيعة القياس التربوى والنفسى :

لكى تتضح طبيعة القياس التربوى والنفسى ربما يكون من المفيد إلقاء مزيد من الضوء على القياس فى العلوم الطبيعية، فالقياس الفيزيائى كما أوضحنا يمكن إجراؤه بطرق مباشرة، حيث إن هناك عمليات وقواعد مقننة متفق عليها لذلك ، كما أن هناك موازين Scales معرفة تعريفًا جيدًا يتم على أساسها تفسير نتائج القياس.

فعند قياس طول منضدة مثلا ما علينا إلا أن نأتى بالمسطرة المدرجة ، ونجعل إحدى حافتى المنضدة مقابلة للنقطة المناظرة لصفر التدريج، ونقرأ الرقم المقابل للحافة الأخرى، وبهذه العمليات البسيطة نستطيع معرفة طول المنضدة بإحدى الوحدات الخطية المعيارية لقياس الطول مثل السنتيمتر أو البوصة.

غير أن هناك ظواهر فيزيائية أكثر تعقيدًا ، من مثل السمع والبصر وما شابههما، تتطلب إجراءات تفصيلية للقواعد أو الطرق التى تتبع فى قياسها. كما أن هناك ظواهر أخرى يصعب قياسها قياسًا مباشرًا،مثل الصوت والضوء والحرارة والكهرباء وغيرها نظرًا لأنها غير مرئية. ولكن هل لابد من رؤية الظاهرة لكي نتمكن من قياسها؟ بالطبع لا، فعلماء الفيزياء استطاعوا قياس هذه الظواهر الطبيعية وغيرها، على الرغم من أنها غير مرئية أو محسوسة، بل إنهم تمكنوا من قياسها بدقة متناهية بواسطة أدوات قياس متنوعة، فنحن لا نقيس الظاهرة في حد ذاتها، وإنما نقيس أثرها فى الأجسام أو الأشياء بواسطة أدوات قياس أعدت لهذا الغرض. فالصوت مثلا يحدث اهتزازات وذبذبات فى الأثير على شكل موجات، وتختلف شدة وتكرار الموجات باختلاف نوع الصوت الصادر من مصدر معين، وأداة القياس أو الجهاز المناسب يقيس هذه التأثرات، ولا يقيس الصوت فى حد ذاته، ولكننا نستخدم المقدار الكمى الناتج عن قياس الأثر فى الاستدلال على ظاهرة الصوت.

وقد حذا علماء النفس والتربية حذو العلوم الطبيعية ، بل إنهم أكثر اهتماما بمنطق القياس والأساليب الكمية من علماء العلوم الطبيعية، ويرجع سبب ذلك إلى أن الظواهر السلوكية ظواهر تتميز بالتعقد والتشابك وتعدد المتغيرات، وليست ببساطة الظواهر الفيزيائية، فعالم التاريخ الطبيعى يستطيع أن يثبِّت زهرة ويقوم بدراستها، وعالم الكيمياء يمكنه أن يتعاملُ مع المواد الكيميائية كما يشاء، وكذلك عالم الفيزياء.

أما عالم النفس فمن الصعب عليه إخضاع الإنسان لتجاربه المعملية، وتثبيت جميع المتغيرات التى يود عزل أثرها.

كما أن كثيرًا من الظواهر السلوكية يصعب قياسها على موازين مدرجة تدريجًا معياريًا متفق عليه كما هو الحال فى موازين العلوم الطبيعية؛ إذ من الصعب قياس سمات شخصية الفرد ، مثل العدوانية أو الانطوائية، أو الإقدام أو التسلطية، أو العاطفة، بنفس القدر من الدقة كما فى حالة قياس درجة حرارة سائل أو شدة تيار كهربائى.

إلا أن هذا لم يقف حائلا دون استمرار علماء القياس التربوى والنفسى فى مواجهة هذه العقبات، وذلك باتباع خطوات مماثلة للخطوات المتبعة فى القياس الفيزيائى، وابتكار طرق وأساليب وأدوات قياس متنوعة تتميز بخصائص محددة سوف نناقشها فيما بعد.

ولتوضيح ذلك نفترض أننا أردنا قياس «معدل القراءة» لدى تلاميذ أحد صفوف مرحلة التعليم الأساسى، فهنا يجب البدء بتعريف المقصود من الخاصة المراد قياسها، وهى فى هذه الحالة «معدل القراءة».

إذ ربما نُعرِّفها بأنها السرعة فى القراءة، أى عدد الكلمات التى يقرأها التلميذ فى الدقيقة. وتصبح المشكلة الآن هى تحديد مجموعة من الإجراءات أو القواعد لقياس سرعة القراءة.

فربما نلجأ إلى أبسط الطرق وهى أن نطلب من كل تلميذ أن يقرأ قطعة نثرية معينة فى مدة زمنية محدودة، ثم نوجد عدد الكلمات التى أمكنه قراءتها فى هذه المدة، ونقسم هذا العدد على الزمن المستغرق لنحصل بذلك على قياس لعدد الكلمات المقروءة فى الدقيقة، وبهذا نكون قد استخدمنا أسلوبًا مماثلاً لأسلوب قياس الطول الذى أشرنا إليه.

ولكننا هنا يجب أن نأخذ بعين الاعتبار مجموعة أخرى من العوامل من بينها صعوبة المقروء، إذ إن هذه الصعوبة تؤثر بلا شك فى معدل القراءة؛ ولذلك فإن عدم معرفتنا درجة صعوبة المقروء يجعل تفسير نتيجة قياس معدل القراءة غير ذى معنى.

والعامل الثانى هو المدة الزمنية المطلوبة للقراءة، فإذا كانت هذه المدة طويلة سوف يشعر التلميذ بالتعب أو الملل ويقل بذلك معدل قراءته. والعامل الثالث هو شروط الموقف الاختبارى، فمثلاً إذا كانت الحجرة غير جيدة الإضاءة أو التهوية أو تحيط بها الضوضاء، فإننا نتوقع أن ينخفض معدل القراءة.

كذلك يجب أن نحدد تحديدا دقيقا المقصود بكلمة «القراءة». وخلاصة ذلك أنه على الرغم من أوجه التشابه بين عمليات وإجراءات القياس الفيزيائى والقياس التربوى والنفسى، إلا أنه يوجد عدد أكبر من المتغيرات التى تؤثر فى القياس التربوى والنفسى يجب ضبطها أو التحكم فيها إذا أردنا أن نجرى عملية القياس بصورة دقيقة.

أنواع القياس ومستوياته :

تبين لنا أن قياس السمات الإنسانية يتطلب تعريفًا دقيقًا لهذه السمات في ضوء المتغيرات المتعلقة بها والعمليات التي نجريها للحصول على قيم لهذه المتغيرات تكون قابلة للتفسير.

ومما لا شك فيه أن الصياغة الإجرائية لتعريف المتغير ليس أمر هينًا، حيث تتطلب مراجعة فاحصة للأطر النظرية والدراسات والبحوث المتعلقة بهذا المتغير، فتعريف المتغير تعريفًا دقيقًا مرتبط ارتباطًا وثيقًا بالعمليات والإجراءات التي نتبعها في قياسه ، وفي كثير من الأحيان نضطر إلى إعادة النظر في تعريف المتغير في ضوء نتائج هذه العمليات .

فالمتغيرات السيكولوجية والتربوية تعد بمثابة تكوينات فرضية Constructs تحتاج بلورتها إلى أدلة متعددة ودراسات متجددة وفقًا لما يطرأ على هذه التكوينات من تغيرات نتيجة تطور المعرفة. فمتغير « التكيف Adjustment» مثلاً مر بهذه المراحل التطورية في تعريفه، فقد كان يستخدم استخدامات متعددة وفقًا لتعريفات متباينة ، مما أدى إلى إرباك ملحوظ لدى الباحثين والممارسين، وقد نتج عن ذلك ضم عدد من السمات أو المتغيرات المنفصلة لتشكل في مجموعها مفهوم التكيف مما تطلب تبريرات فكرية ومنطقية بل وإحصائية لهذا الإجراء.

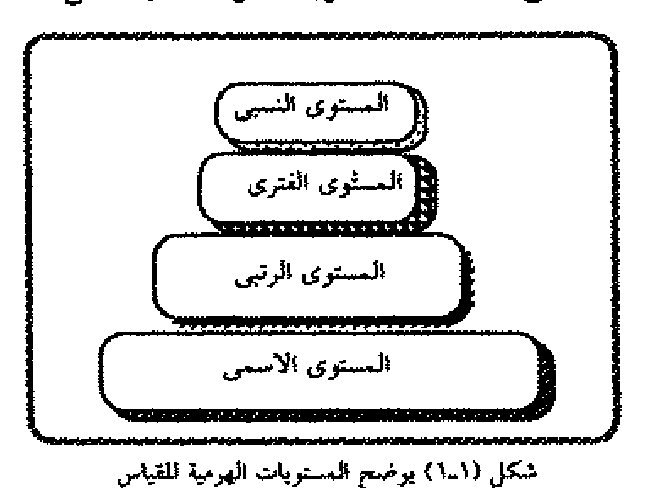
كما أن متغير «التصلب Rigidity» مثلا لدى الأفراد تباينت بل وتناقضت نتائج البحوث التي اهتمت بدراسته مما أدى إلى مزيد من الغموض حول طبيعة هذا المتغير ، وما إذا كان له وجود فعلي.

لذلك فإن من الضروري العناية بتعريف المتغير المراد قياسه ليكون أساسًا لتصميم سلسلة من القواعد والإجراءات التي تكشف عن الخاصة أو السمة المتعلقة به وقياسها.

ومن الجدير بالذكر أن المتغيرات النفسية والتربوية إما أن تكون كيفية أو كمية ، فالمتغير الكيفي أو النوعي تناسبه عمليات وإجراءات التصنيف، ويطلق عليه حينئذ المتغير الاسمي، أما المتغير الكمي فتناسبه عمليات وإجراءات تسمح بالحصول على ترتيب تصاعدي أو تنازلي لقيمه، وعندئذ يطلق عليه المتغير الرتبي،أو تسمح بالحصول على قياسات كمية له، ويطلق عليه المتغير الفتري أو النسبي.

لذلك تتباين أنواع القياس ومستوياته وفقا لنوع المتغير وطبيعته والهدف من عملية القياس ، فلكي نجري عملية القياس بالدقة المطلوبة يجب أن نراعي مستوى قياس المتغير. وتستند كل من هذه المستويات إلى فروض رياضية ومنطقية، وتتخذ تنظيما

هرميًا تحقق فيه المستويات الأعلى فروض جميع المستويات الأدنى بالإضافة إلى خصائص تميز المستوى الأعلى ، وهذه المستويات موضحة بالشكل التخطيطى التالى:



شكل (١ـ١) يوضح المستويات الهرمية للقياس

ويتضح من شكل (١ـ١) أن هناك أربعة مستويات للقياس وفقا لطبيعة المتغيرات المراد قياسها، وهذه المستويات هى :

المستوى الاسمى Nominal Scale :

وهو أدنى مستويات القياس ويناسب المتغيرات الكيفية أو النوعية التى تتطلب تصنيف الأفراد إلى مجموعات منفصلة للتمييز بينهم فى سمة معينة، ويكون الهدف من عملية القياس فى هذه الحالة هو التصنيف Classification الذى يراعى الفروق النوعية بين الأفراد. والأعداد المستخدمة فى هذا المستوى من القياس تعد بمثابة رموز بسيطة تستخدم كأسماء لفئات أو مجموعات منفصلة ومتمايزة. ومن أمثلة متغيرات هذا المستوى : النوع، والجنسية، والديانة، والحالة الاجتماعية، والانتماء إلى مؤسسات معينة، وهكذا.

وهذه الأعداد لا نستطيع إجراء عمليات حسابية عليها بحيث تكون ذات معنى، فلا معنى لأن نجمع رقم مناظر لنوع معين على رقم مناظر لنوع آخر مثل الذكور والإناث، أو نطرح رقم مناظر لجنسية معينة من الرقم المناظر لجنسية أخرى، ومع هذا يمكن التعامل مع البيانات الناتجة عن القياس الاسمى بأساليب إحصائية مناسبة، سوف نشير إلى البعض منها فى الفصل الثالث.

المستوى الرتبى Ordinal Scale :

وفى هذا المستوى يمكن ترتيب الأفراد أو الأشياء تبعًا لخاصة أو سمة معينة. فمثلاً يمكن أن نطلب من الأفراد ترتيب ميولهم المهنية، إذ ربما يعطى فرد مهنة مبرمج

الحاسب الآلي الرقم ١ ، ويعطى مهنة مهندس الرقم ٢، وفنى سيارات الرقم ٣ وهكذا، وهذا يعنى أن هذا الفرد قد أجرى نوع من القياس الرتبي لميوله المهنية وفقًا لمحكات معينة استند إليها في هذا الشأن.

ونظرًا لسهولة ويسر هذا النوع من القياس ، فإنه يستخدم بكثرة في القياس التربوى والنفسى وبخاصة في قياس السمات الوجدانية التي تتعلق بالاتجاهات والتفضيلات والآراء والتقديرات. ولكن ينبغى ملاحظة أنه ليس من الضرورى معرفة مقادير أو درجات السمات المقاسة لكى يمكننا إجراء هذا الترتيب، بل يكفى معرفة ترتيب كل منها بالنسبة للآخر. كما أن القياس الرتبي لا يفترض أن الفروق بين الرتب تساوى الفروق بين درجات السمة المقاسة.

وهذه الأعداد التي تمثل الرتب لا تسمح بإجراء عمليات حسابية عليها شأنها في ذلك شأن القياس الاسمى ، وإنما نكتفى بالقول بأن محمدًا أذكى من أشرف ، أو تحصيل محمد أعلى من تحصيل أشرف، أو أن محمداً لديه اتجاه نحو المدرسة أكثر إيجابية من أشرف ، وهكذا.

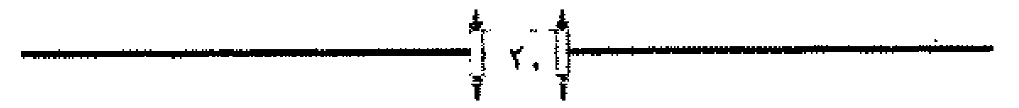
المستوى الفترى Interval Scale :

وهذا المستوى من القياس يتعلق بتحديد مقدار الفرق بين شيئين، والمثال الشائع لذلك هو الترمومتر، فبعض الترمومترات تكون تداريجها درجات فهرنهيتية، حيث يمثل العدد ٣٢ درجة فهرنهيتية درجة التجمد ، والعدد ٢١٢ درجة فهرنهيتية درجة الغليان، والأعداد المحصورة بينهما تشير اعتباريًا إلى تمدد أو تقلص معين في عمود الزئبق أو السائل في الترمومتر، فتغير قدر معين من هذا السائل وليكن ١,٢ ملليمتر تعنى تغير في الحرارة مقداره درجة واحدة، بغض النظر عما إذا كان هذا الفرق بين الدرجتين ٥، ٦ أو بين ٦٢، ٦٣ درجة فهرنهيتية.

فالفروق المتساوية على طول التدريج تمثل فروقًا متساوية في حركة السائل داخل الترمومتر.

وهنا ينبغى ملاحظة أن الصفر الحقيقي الذى عنده تنعدم الحرارة لا وجود له على مثل هذه الترمومترات، فالترمومتر يقيس فقط الفرق الحقيقي بين أى درجتين من درجات الحرارة.

وهذا المستوى من القياس يستخدم كثيراً في القياس التربوى والنفسى، فنحن لا نقيس ذكاء الفرد أو سمات شخصيته أو ميوله قياسًا مطلقًا، وإنما نقيس الفرق الحقيقي بين ذكاء شخصين مثلاً، أونوازن بين ذكائه وذكاء زملائه الذين يطبق عليهم نفس



الاختبار، وتمثل الفروق المتساوية على متصل تدريج السمة (الذكاء) فروقًا متساوية فى الذكاء بالفعل.

ولكن لا يوجد صفر مطلق للذكاء أو لأى سمة أخرى من سمات شخصية الإنسان، فالسمات لا تنعدم لدى الأفراد.

المستوى النسبي Ratio Scale :

يتميز هذا المستوى من القياس بالخصائص التى تتوافر فى المستوى الفترى بالإضافة إلى وجود صفر مطلق على ميزان القياس يناظر بالفعل انعدام الخاصة أو السمة المقاسة، ومثال ذلك الأطوال والكتل وغير ذلك. ونظرًا لأنه يمكن إجراء جميع العمليات الحسابية الأساسية فى هذا المستوى فإنه يعد أعلى المستويات السابقة. ويستخدم هذا النوع من القياس عادة فى العلوم الطبيعية، ويندر استخدامه فى العلوم السلوكية للأسباب السابقة، فمعظم الاختبارات والمقاييس التربوية والنفسية تؤدى عادة إلى قياسات فترية، ولكن ينبغى مراعاة العمليات والإجراءات التى سبق أن أشرنا إليها.

مدخل السمات فى القياس التربوى والنفسى :

ذكرنا فيما سبق أن تحديد السمات المراد قياسها وتعريفها تعد من الخطوات الأساسية فى القياس. فالقياس التربوى والنفسى لا يهدف لقياس الفرد، وإنما قياس خصائصه أو سماته أو مظهر معين من مظاهره كالذكاء أو دافعية الإنجاز، أو القلق مثلاً. لذلك فإن مفهوم السمة Trait يُعد من المفاهيم التى تحتاج إلى مزيد من التوضيح. فقد وجد بعض الباحثين أن هناك ما يزيد عن ٤٠٠٠٠ كلمة تصف الأفراد. وبالطبع يصعب استخدام هذا القدر من الكلمات أو التعامل معها في القياس التربوى أو النفسى؛ لذلك حاول علماء النفس اختزال هذا الكم الهائل من الأوصاف إلى مجموعة أقل من السمات فى إطار ما يعرف بنظرية السمات Trait Theory.

والسمة مفهوم يستخدم لوصف سلوك الأفراد، وهى عبارة عن تجمع من السلوك المترابط، أى الذى يحتمل حدوثه معًا. فالسمة بهذا المعنى ليست صفة منفردة، وإنما تعد مفهومًا مجردًا، وليس شيئًا ملموسًا، ويمكن أن تكون السمة عامة مثل سمة «الميل إلى المشاركة فى مواقف اجتماعية متنوعة»، أو تكون محدودة مثل سمة «تذوق نوع معين من الطعام».

والسمات لا يمكن قياسها قياسًا مباشرًا، وإنما يستدل عليها من أنماط السلوك الملاحظ، وهذا هو الحال دائمًا فى القياس التربوى والنفسى. فعندما يود عالم النفس

قياس إحدى السمات العقلية، ولتكن الذكاء، فإنه عندما يلاحظ طفل معين لا يرى ذكاؤه لأن الذكاء مفهوم مجرد، ولكنه يلاحظ سلوك الطفل في مواقف مختلفة تتطلب الذكاء، أي أن السمات العقلية تعد بمثابة خصائص يفترض أنها تعكس مجموعة مترابطة من السلوك الذي يمكن ملاحظته وتسجيله في مواقف اختبارية مقننة . ويجب أن تفي السمات ببعض الشروط لكي يتسنى الإفادة منها في وصف السلوك، ومن أهمها:

(١) أن تمثل السمة خاصة محددة يتباين فيها الأفراد مثل السمات العقلية والشخصية والوجدانية .

(٢) أن تكون السمة قابلة للتعريف الإجرائي بحيث يكون هناك قدر معقول من الاتفاق بين القائمين بملاحظة السلوك حول مقدار السمة لدى الأفراد المختلفين، فمثلاً ما الذي نعنيه بسمة «الذكاء» ؟ وما نوع السلوك الذي يميز الشخص الذكي أو يسمه؟ وهل نقصد بالذكاء تناول الأفكار والمفاهيم المجردة ، أم التعامل مع المواقف التي تتميز بالجدة، أم استجابات الفرد في مواقف مألوفة ، أم سرعة الاستجابة وطلاقتها، أم إلى مستوى تعقد الاستجابة بغض النظر عن زمنها؟

لذلك ينبغي التوصل إلى تعريف واضح ودقيق ومقبول للسمة المراد قياسها، كما ينبغي أن تتضمن الجوانب المهمة التي نود التركيز عليها من حيث فائدتها العملية أو التطبيقية.

(٣) أن تتميز السمة بالثبات النسبي بمرور الزمن وباختلاف المواقف، فسلوك الفرد يصعب التنبؤ به من موقف إلى آخر، ومن وقت إلى آخر مالم يكن متميزاً بالاتساق. ومن الجدير بالذكر أن السمات الإنسانية تتباين في مدى ثباتها واتساقها ، فالسمات العقلية تتميز بالثبات النسبي بدرجة أكبر من السمات الشخصية أو الوجدانية.

بعض الخصائص الأساسية للسمات :

أشار كيلي (Kelly , 1974) ، وثورنديك (Thorndike ,1974) ، ونانلي (Nunnally, 1998) إلى مجموعة من الخصائص الأساسية للسمات التي تيسر عمليات قياس السلوك الذي يستدل منه على سمة معينة، وفيما يلي هذه الخصائص:

(١) يمكن أن تمثل السمة بخط متصل : إن الأخذ بمفهوم السمة الذي أشرنا إليه يجعلنا نفترض أن هناك متصلاً من السلوك نحاول عن طريق عملية القياس أن نحدد موقع الفرد عليه في سمة معينة لديه. ونقصد بالمتصل Continuum خط مستقيم

يتكون من عدد لا نهائى من النقط الممكنة التى تحدد مواقع مختلفة للسمة المقاسة. فالذكاء مثلاً ليس سمة ثنائية القطب، أى تمثل بقسمين منفصلين هما الذكى والغبى، وإنما تمثل بمقادير متصلة بحيث يمكن تتبع سمة الذكاء فى مستويات مختلفة متدرجة يتمايز فيها الأفراد بحسب درجاتهم فى هذه السمة.

(٢) يمكن تدريج متصل السمة بوحدات مناسبة: إن متصل السمة يمكن تدريجه بحيث يصبح بمثابة ميزان Scale ، ويعتمد هذا التدريج على طبيعة السمة المراد قياسها، والفروض التى نستند إليها فى قياسها، وتعريفنا لوحدات الميزان. إذ ربما نقتصر على الميزان الاسمى الذى يصنف الأفراد إلى مجموعات منفصلة فى سمة معينة مثل انطوائى وانبساطى، وربما نستخدم الميزان الرتبى الذى يصنف الأفراد مع مراعاة الترتيب مثل عبقرى، وذكى جداً، وفوق المتوسط، ومتوسط الذكاء، ودون المتوسط، أو ربما يفضل الميزان الفترى الذى يراعى تساوى الفروق بين الأقسام المتتالية على متصل السمة المقاسة، أى تساوى الوحدات.

(٣) يمكن تصور متصل السمات على أنه ميزان قياس فترى : فعلى الرغم من تعدد موازين القياس، إلا أنه يمكن جعل ميزان كثير من السمات النفسية من المستوى الفترى. فالقياس فى واقع الأمر ـ كما سبق أن ذكرنا ـ هو محاولة لإيجاد تناظر بين الأعداد وبين خصائص الظاهرة موضع القياس، فكلما كان التناظر كبيراً دل ذلك على أن القياس مناسب. ويؤدى ذلك إلى اعتبار مهم هو ما إذا كان مستوى معين من مستويات القياس يناظر بدرجة أكبر واقع الظاهرة عن غيره من المستويات. وينصب عادة اهتمامنا فى القياس على ما إذا كان المستوى الفترى يمثل واقع الظواهر النفسية والتربوية تمثيلاً مناسبًا وكافيًا.

ولتوضيح ذلك نفترض أننا نود قياس القدرة الحسابية لدى الطلاب، فإحدى طرق قياس هذه القدرة هو أن نقوم ببناء اختبار جيد يشتمل على العمليات والمهارات الحسابية، وتكون الدرجة التى يحصل عليها كل طالب بمثابة مؤشر لقدرته الحسابية. فإذا افترضنا أن الاختبار يشتمل على ٥٠ مفردة تختلف فى صعوبتها ، فهل يمكننا اعتبار أن فرق خمس درجات على جزء من متصل ميزان الدرجات يكافئ فرق خمس درجات على أى جزء آخر من الميزان، كما هو الحال فى درجات الحرارة مثلا؟ أى هل الفرق بين الدرجتين ٥ ، ١٠ مثلا على ميزان الدرجات يكافئ الفرق بين ٤٥ ، ٥٠ على الميزان نفسه؟

وهنا ربما نفترض أن كل نقطة أو درجة (أى كل مفردة أجاب عنها الطالب إجابة صحيحة) تساوى أى نقطة أو درجة أخرى يحصل عليها في الاختبار، وبذلك نعتبر الفرق بين الدرجتين ٥ ، ١٠ مكافئًا للفرق بين الدرجتين ٤٥ ، ٥٠ في مقدار السمة المقاسة وهي القدرة الحسابية. ولكن إذا تمعنا قليلاً ربما نستنتج أن الفرق بين الدرجتين ٤٥ ، ٥٠ أكبر من الفرق بين الدرجتين ١٠ ، ١٥ ، إذ إنه لكي يرفع الطالب درجته من ٤٥ إلى ٥٠ ربما يتطلب الإجابة عن خمس مفردات أخرى أكثر صعوبة، بينما يتطلب الأمر الإجابة عن خمس مفردات أقل نسبيًا في صعوبتها إذا أراد أن يرفع درجته من ١٠ إلى ١٥ ، فإذا قبلنا هذا التبرير كان معنى ذلك أن ميزان الدرجات ليس من المستوى الفترى.

ويرى كرونباك Cronbach أن الاختلاف بين الافتراضين أو التبريرين يعتمد على الأسلوب الذى نتبعه في تفسير درجات الاختبار وتبريرنا لهذا التفسير. فالتبرير الأول المتعلق بتساوى الفروق بين الدرجات يمكن اعتباره تبريراً منطقيًا، أما التبرير الثانى فهو تبرير سيكولوجى حيث يأخذ بعين الاعتبار واقع السمة موضع القياس، أى يهتم بمدى تأثر التحسن فى الأداء باختلاف مدى الدرجات وفقا لصعوبة أو سهولة مفردات الاختبار، حيث إن الصعوبة أو السهولة لها بالطبع جانب نفسى.

وهنا ربما يتساءل القارئ: كيف إذن يمكن أن نحصل على ميزان فترى أو نسبى فى القياس التربوى والنفسى؟ وللإجابة عن هذا التساؤل ينبغى مراعاة متطلبات الميزان النسبى التى أشرنا إليها من قبل، وهى الفترات المتساوية وتوافر صفر مطلق يناظر انعدام السمة. ونظراً لأنه يصعب تصور إنسان ذكاؤه أو قلقه أو تحصيله صفرًا، فإنه لا يوجد صفر مطلق على ميزان قياس السمات الإنسانية، ولكن يمكن الاستعاضة عنه بصفر اعتبارى مثل درجة الحرارة صفر. وهذا الصفر لا يناظر انعدام السمة، وإنما يمكن اعتباره نقطة تقع فى منتصف الخط المتصل الذى يمثل السمة، أى أن الصفر يناظر الموقع المحايد على ميزان قياس هذه السمة، كما أن عدد الوحدات أو التداريج التى يقسم إليها الميزان تكون اعتبارية أيضًا، ويمكن جعلها متساوية باستخدام بعض الأساليب الإحصائية التى سنوضحها فى الفصل الثالث.

وهذا الإجراء يمكننا من قياس السمات الإنسانية على ميزان فترى بدلا من الاقتصار على الميزان الرتبى المميز لدرجات الاختبارات فى كثير من الأحيان. لذلك يهتم علماء القياس بالحصول على ميزان فترى فى المقاييس التربوية والنفسية، وذلك لإمكانية تحويل الدرجات الناتجة عن هذه المقاييس تحويلاً خطيًا يسمح بإضافة مقدار

ثابت إلى الدرجات أو طرحه منها دون أن تتأثر العلاقة بين الدرجات، وبهذا نستطيع تحويل الدرجات من ميزان ذى وحدات معينة إلى ميزان له وحدات مختلفة، أى يمكن تحويل الدرجات إلى ميزان مشترك للموازنة بين أداء الأفراد على مقاييس مختلفة. ولعل درجات الحرارة تعد مثالاً جيداً لذلك، فعلى الرغم من أننا نستطيع قياس الحرارة بكل من وحدات الدرجات المئوية، والدرجات الفهرنهيتية، إلا أنه يمكن تحويل كل من الوحدتين إلى الأخرى ووضعهما على ميزان مشترك.

أنواع السمات النفسية والتربوية:

نظراً لتنوع وتعدد السمات النفسية والتربوية، فقد بذلت محاولات كثيرة لتصنيف هذه السمات لتيسير دراستها وقياسها والإفادة منها فى فهم وتفسير السلوك الإنسانى، إلا أن هذه المحاولات لم تتوصل إلى اتفاق عام فى هذا الشأن.

فالتساؤل الذى يتردد بين الباحثين والممارسين فى مختلف مجالات القياس والتقويم هو : ماذا نقيس؟

وعلى الرغم من أهمية وضرورة هذا التساؤل، إلا أن الإجابة عنه تتطلب معرفة الهدف من القياس، وطبيعة البحث أو الدراسة، ونوع القرارات العملية التى يجب اتخاذها فيما يتعلق بالأفراد الذين نود قياس بعض سماتهم سواء فى المدارس أو الجامعات أو المستشفيات أو فى مجال الصناعة أو المؤسسات العسكرية.

فنحن بالطبع لا نستطيع قياس الفرد ككل، وإنما نهتم بقياس بعض سماته التى تسهم فى اتخاذ قرار معين بشأنه يتعلق بالانتقاء أو التصنيف أو التسكين أو التشخيص مثلا. وقد ميز كيلى(Kelly, 1974)بين عدة مجموعات من السمات التى تهم الباحثين والمشتغلين بالقياس والتقويم التربوى والنفسى، بعضها يتعلق بالجوانب الجسمية العضوية والفسيولوجية، والبعض الآخر يتعلق بالجوانب النفسية العقلية والوجدانية والمزاجية والمهارية كما يلى:

(١) سمات تتعلق بالشكل الخارجى لجسم الفرد: وهذه السمات يغلب عليها الطابع الجسمى العضوى وليس الطابع النفسى، وإن كانا أحياناً مترابطين. فالسمات التى تميز جسم الفرد مثل الطول والوزن وحجم الكفين والقدمين وشكل الصدر مثلا ربما ترتبط بجوانب أخرى فى الشخصية وبخاصة مفهوم الفرد عن ذاته. ومع هذا يرى بعض علماء النفس أن هذه السمات لا تنتمى إلى مجال الشخصية الإنسانية، ولكنها بلاشك خصائص يتباين فيها الأفراد تبايناً ملحوظاً.

(٢) سمات فسيولوجية : وتتعلق بالسلوك الداخلى لأعضاء الجسم الحيوية، مثل ضربات القلب، وضغط الدم، ودرجة حرارة الجسم، ونشاط الغدد الدرقية، وغير ذلك. ويتباين الأفراد أيضًا تباينًا ملحوظًا فى هذه السمات. وترجع أهمية هذه السمات إلى أنها تؤثر فى الجوانب النفسية للفرد وتتأثر بها. فالتغيرات فى القياسات الفسيولوجية تتباين تباينًا كبيرًا بتباين المواقف الحياتية والضغوط الفيزيائية والنفسية؛ لذلك تستخدم مقاييس مختلفة للأنشطة الفسيولوجية فى تقييم الشخصية.

(٣) الاستعدادات: وهى القدرة الكامنة لدى الأفراد أو قابليتهم لأداء عمل معين. ويتباين الأفراد تباينًا كبيرًا فى هذه الاستعدادات حتى قبل أن تتاح لهم الفرصة للتحصيل أو اكتساب مهارات معينة، ونظرًا لاتساع مجال هذه المهارات وما تتطلبه من استعدادات مختلفة، فإنه يصعب حصر هذه الاستعدادات حصرًا دقيقًا، ومن بين هذه الاستعدادات:

الاستعداد الدراسي: ويتعلق بالقدرات اللازمة للنجاح فى المدرسة وبخاصة القدرات اللفظية والقدرات الرياضية.

الاستعداد المكاني: ويتعلق بالقدرة على تصور الأشكال ودورانها فى المستوى أو فى الفراغ.

الاستعداد الميكانيكي: وهو قدرة الفرد على اكتساب المعارف والمهارات المتعلقة بفك وتركيب أجزاء الأشياء، والمهارات اليدوية فى التعامل مع الأجزاء الدقيقة المكونة للأدوات المختلفة.

الاستعداد لأداء الرياضات: وهو القدرة على أداء لعبة رياضية معينة أو أكثر بدقة ومهارة، ويتطلب ذلك بنية جسمية قوية، وقدرة إدراكية، وقوة عضلية وتناسق.

الاستعداد الموسيقي: وهو القدرة على تعلم العزف على آلة موسيقية أو أكثر بيسر وسهولة، وكذلك مهارة تأليف الألحان.

الاستعدادات الفنية الأخرى: وتتعلق بالقدرة على الإنتاج الفنى المبدع، مثل الغناء والتمثيل، وتذوق إنتاج الآخرين.

الاستعداد الكتابي: وهو القدرة على تعلم المهارات الكتابية، مثل استخدام الحاسوب ، والكتابة على الآلة الكاتبة ، والاختزال ، وأعمال الأرشيف أو حفظ المستندات والوثائق.

(٤) المهارات والتحصيل: وهذه تتعلق بالأداء الفعلى للأفراد سواء فى المدرسة أو فى مواقع العمل المختلفة التى تتطلب مهارات معينة. وكما أن هناك تباينًا بين الأفراد فى استعداداتهم وإمكاناتهم، كذلك يوجد تباين بينهم فى المهارات التى تتطلب

توفر قدرًا معينًا من الاستعدادات المتعلقة بهذه المهارات. وعلى الرغم من أن هذه الاستعدادات تستخدم كمنبئات للسلوك المستقبلي، وتُيسر اكتساب المعارف والمهارات اللازمة لأداء عمل معين، إلا أن الأداء الفعلي مشروط أيضًا بالدافعية والحافز والممارسة والدراسة، أي أن الفروق بين الأفراد في الأداء والتحصيل يمكن أن تعزى إلى الأثر المشترك لكل من الاستعدادات، وهذه المتغيرات الأخرى.

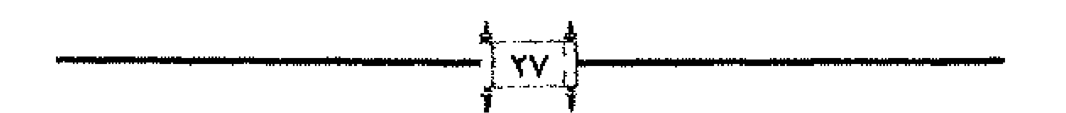
(٥) **الميول:** تتعلق بالجوانب الوجدانية لشخصية الفرد، وينعكس ميله في تعبيره عما يحب ويكره، والأنشطة التي يفضل أن يبذل فيها جهدًا ووقتًا، مثل قراءة الكتب العلمية، والعناية بالنباتات، وسماع الموسيقى، وغير ذلك.

(٦) **الاتجاهات:** تعد أيضًا من الجوانب الوجدانية، وتتمثل في آراء الفرد وتفضيلاته فيما يتعلق بجماعات معينة أو مؤسسات أو ممارسات أو سياسات أو مفاهيم، وما شابه ذلك. وينعكس اتجاه الفرد نحو القضايا المختلفة في تعبيره وسلوكه، ومن أمثلة ذلك: الاتجاه نحو حقوق الإنسان، أو تنظيم الأسرة، أو نظام الضرائب، أو أماكن العبادة، وهكذا.

(٧) **القيم:** وهي أيضًا من الجوانب الوجدانية، ولكنها أكثر تأصيلاً وعمقًا من الميول والاتجاهات، وتتعلق بفئات واسعة من الخبرات والمعتقدات والأهداف والسلوكيات التي تتميز بالاستغراق والولاء، فالفنان لديه قيم جمالية، ورجل الدين لديه قيم دينية، ورجل الأعمال لديه قيم اقتصادية، وهكذا.

(٨) **السمات المزاجية:** وهي تعد من سمات الشخصية التي نالت قدرًا كبيرا من اهتمام علماء النفس؛ نظرًا لأنها من السمات التي تسهم في تكيف الفرد وتفاعله مع الآخرين، وهي تعبر عن نزعات الفرد أو طباعه مثل التفاؤل، وحدة المزاج، والثقة بالذات، والاندفاع، والانبساط، والانضباط، والإحساس بالأمان، وغير ذلك من النزعات التي تميز الفرد في مواقف المشاركة الاجتماعية والمواقف الانفعالية وغير الانفعالية.

والخلاصة أن هذه المجموعات من السمات يتباين الأفراد فيها تباينًا ملحوظًا. وعلى الرغم من أنها لا تعد المجموعات الوحيدة، إلا أنها تفيد في قياس السلوك الإنساني وتقييمه. وبالطبع يصعب قياسها جميعًا لدى فرد واحد، ولكن ينبغي انتقاء السمات المتعلقة بالهدف من القياس، وبالاستخدامات المناسبة لنتائج القياس. فلكل فرد استعداداته وميوله، وخصائص الشخصية التي تؤهله للقيام بدور فاعل في مجتمعه، وتحديدنا لهذه السمات وقياسها لديه يمكننا من الإفادة القصوى من إمكانات كل فرد وتنظيم دور كل منهم في الحياة بما يتفق وهذه الإمكانات.



عمليات القياس وأدوات القياس :

ينبغى أن نفرق بين عمليات القياس التى أوضحناها فيما سبق وأدوات القياس، أى الاختبارات والمقاييس. فأدوات القياس تستخدم فى جمع بيانات بطريقة منظمة تفيد فى صنع القرارات المختلفة، وهذه الأدوات تشتمل على مطلب أو مجموعة من المطالب أو المهام أو المثيرات التى ترتبط بسمة نفسية أو تربوية.

وعادة تتطلب أدوات القياس من الفرد أن يستجيب لهذا المطلب أو المطالب (مفردات أو أسئلة أو فقرات) بحيث يمكن الاستدلال من إستجابته على مقدار السمة المراد قياسها. وبالطبع لا تقتصر أدوات القياس على الاختبارات التى تقيس الجوانب العقلية والتحصيلية، وإنما تشمل الاستبيانات وقوائم الملاحظة والتقارير الذاتية والطرق الإسقاطية وغيرها من أدوات قياس مجموعات السمات السابقة. وتفيد هذه الأدوات فى جمع بيانات من مصادر متعددة للاستفادة بها فى عمليات التقويم. وسوف نلقى الضوء فيما يلى على مفهوم الاختبارات، ومفهومى التقويم والتقييم وعلاقتهما بمفهوم القياس لكى يتسنى توظيف هذه المفاهيم توظيفًا مناسبًا.

الإختبار التربوى والنفسى :

إن تعدد السمات واختلاف طبيعتها وتباين أنواعها يتطلب أن تتباين أدوات قياس أو تقدير هذه السمات كما أشرنا، لذلك ربما لا نجد تعريفًا مناسبًا ينطبق على جميع هذه الأدوات. فمثلاً يعرِّف ساكس (Sax,1974) الاختبار النفسى والتربوى بأنه مطلب أو مجموعة من المطالب تستخدم للحصول على ملاحظات منظمة يفترض أنها تمثل سمة نفسية أو تربوية، ويعرِّفه نيتكو (Nitko 1983) بأنه إجراء منظم لملاحظة ووصف سمة أو أكثر من سمات الفرد بالاستعانة بميزان قياس أو نظام تصنيف معين. أما براون (Brown1976) فيعرِّفه بأنه إجراء منظم لقياس عينة من السلوك. غير أن التعريف الذى اقترحه شيس (Chase 1985) على الرغم من أنه يتفق مع تعريف نيتكو (Nitko) إلا أنه أكثر عمومية لاشتماله على المكونات الأساسية المشتركة بين أدوات القياس. فقد اقترح شيس (Chase) أن الاختبار النفسى والتربوى هو أداة قياس مقننة أو أسلوب منظم يصمم للحصول على قياس موضوعى لعينة من السلوك بهدف موازنة أداء الفرد بمعيار أو بمستوى أداء محدد.

وهذا التعريف يعد تعريفًا شاملاً للأسلوب الذى يتبع فى بناء الاختبار، وطريقة تسجيل الملاحظات عن السلوك، وتصحيح أو تقدير درجات الاختبار. ولعل من

المناسب إلقاء الضوء على المكونات الأساسية التى يشتمل عليها هذا التعريف،وهى : التقنين، والموضوعية، وعينة السلوك، ومعيار أو محك الأداء.

التقنين Standaradization: ويقصد به أن يكون بناء وتصحيح وتفسير نتائج الاختبار أو أداة القياس مستنداً إلى قواعد محددة بحيث تتوحد فيه وتتحدد بدقة مواد الاختبار وطريقة تطبيقه ، وتعليمات إجابته وطريقة تصحيحه أو تسجيل درجاته، وبذلك يصبح الموقف الاختبارى موحداً بقدر الإمكان لجميع الأفراد فى مختلف الظروف. والأسلوب الموضوعى المنظم فى بناء الاختبار يعنى أن يتم اختيار مفرداته أو كتابتها بطريقة منظمة تتفق مع مواصفات محددة للاختبار، وأن تمثل هذه المفردات السمة المراد قياسها.

ومفهوم التقنين ينطبق على الظروف المثالية التى يستطيع فيها مصمم الاختبار أو أداة القياس ضبط جميع المتغيرات ما عدا المتغير الذى يهدف الاختبار لقياسه، وبالطبع يصعب عمليًا تحقيق ذلك بهذا القدر من الدقة.

ولعل مفهوم التقنين يتعلق بما سبق أن أشرنا إليه فى مستهل هذا الفصل وهو أن البحث العلمى يتميز بقابلية التكرار وتداول النتائج وعموميتها، فإذا أردنا مقارنة النتائج التى يحصل عليها مختلف الباحثين يجب أن تكون أدوات القياس وتعليمات استخدامها موحدة للجميع لكى يتسنى جمع الملاحظات العلمية فى ظروف مضبوطة، وهو ما يُعرف فى البحث العلمى بالضبط التجريبى.

فالاختبار أو أداة القياس التربوى والنفسى يمكن تمثيلها بتجربة علمية مصغرة تُضبط فيها جميع جوانب الموقف الاختبارى واستبعاد العوامل المختربة لجميع الأفراد، فإذا ما انتظمت هذه الشروط، فإنه يمكن أن نستنتج بقدر كبير من الثقة أن الدرجة التى يحصل عليها الفرد فى الاختبار تمثّل مقدار السمة المقاسة لديه.

الموضوعية Objectivity : ويُقصد بها ألا يتأثر تطبيق الاختبار أو أداة القياس وتصحيحه وتفسير نتائجه بالحكم الذاتى للقائمين بهذا العمل. وتختلف درجة الموضوعية التى يمكن تحقيقها، فالموضوعية التامة تعد هدفًا يسعى إلى تحقيقه القائمون ببناء الاختبارات وأدوات القياس. ويمكن توضيح ذلك بالاختبار الذى يشتمل على مفردات اختيار من متعدد حيث يطلب من الفرد اختيار إجابة واحدة صحيحة من بين أربع أو خمس إجابات معطاة، ففى هذه الحالة يمكن تطبيق مفتاح تصحيح معين على أوراق الإجابة؛ وبذلك لا تختلف الدرجة باختلاف الشخص القائم بعملية التصحيح.

أما الاختبارات أو أدوات القياس التي يعتمد تصحيحها على الحكم الذاتي للشخص القائم بعملية التصحيح، فإننا لا نتوقع أن يكون هناك اتفاق تام بين القائمين بذلك.

ومن الجدير بالذكر أن مفهوم الموضوعية في القياس اتخذ اتجاهًا معاصرًا مختلفًا سوف نوضحه في الفصل الرابع عشر.

عينة السلوك Sample of Behavior : فالاختبار أو أداة القياس تشتمل عادة على مجموعة من المفردات التي تقيس عينة من السلوك الذي نستدل منه على السمة المراد قياسها؛ إذ لا يعقل أن يشتمل الاختبار على جميع المفردات التي يحتمل أن تقيس سمة معينة لدى الأفراد؛ ولذلك يجب مراعاة أن تكون عينة مفردات الاختبار ممثلة للسلوك المراد قياسه ولجميع المفردات التي يمكن أن تقيس هذا السلوك. ويعتمد حجم هذه العينة على طبيعة ومكونات السمة المراد قياسها، والزمن المتاح للإجابة عن مفردات الاختبار، وعلى الكلفة، وغير ذلك من العوامل. ومشكلة تحديد عينة السلوك تعد من مشكلات صدق الاختبار Test Validity .

وهناك مشكلة أخرى ترتبط بمشكلة الصدق، وهي مدى اتساق الدرجة التي يحصل عليها الفرد في اختبارين يشتمل كل منهما على عينة مختلفة من المفردات الممثلة للسمة المقاسة، وتعد هذه من مشكلات ثبات درجات الاختبار Test Reliability. ونظرًا لأن هاتين المشكلتين تتعلقان بخصائص الاختبارات التربوية والنفسية، فسوف نتناولهما بإسهاب في الفصلين الرابع والخامس.

معيار أو محك الأداء : فدرجة الفرد في اختبار ما أو أداة قياس معينة يصعب تفسيرها دون تحديد معيار أو محك أداء معين، فإذا حصل طالب في اختبار ما على الدرجة ١٣٠، فإن هذه الدرجة ليس لها معنى في ذاتها، وإنما تستمد معناها من معيار Norm جماعة معينة، وليكن متوسط درجات الفصل الذي ينتمي إليه الطالب، أو من محك Criterion أداء مقبول. فمثلاً إذا طبق الاختبار على طلاب الفصل وكان متوسط درجاتهم ١٠٠، وأعلى الدرجات ١٢٠ وأدناها ٥٠، فإن الدرجة ١٣٠ التي حصل عليها الطالب يصبح معناها أكثر وضوحًا، أي أن مقارنة أداء الطالب بأداء أقرانه يجعل للدرجة التي يحصل عليها معنى ودلالة؛ لأنها عندئذ تحدد مركزه النسبي بين أقرانه، وبهذا يكون المعيار الذي تفسر الدرجة في ضوئه هو معيار الجماعة، وبعبارة أخرى نقول: أن التفسير مرجعي الجماعة أو المعيار Norm - Referenced .

غير أن هذا التفسير لا يفيدنا فى معرفة مدى إتقان الطالب للمهارات أو المهام التى يقيسها الاختبار، إذ ربما تكون الدرجة ١٣٠ التى حصل عليها الطالب أقل من الحد الأدنى المطلوب للنجاح فى أداء المهارات المرجوة على الرغم مما يبدو من مركزه النسبى المرتفع، فالعبرة هنا ليست بمقارنة الدرجة بمعيار جماعته أو أقرانه، وإنما بمقارنتها بمحك الأداء المتعلق بالمهارات أو المهام المراد قياس مدى تمكن الطالب منها بغض النظر عن ترتيبه فى الفصل. وبهذا تفسر الدرجة فى ضوء محك أداء محدد، أى أن التفسير يكون مرجعى المحك Criterion - Referenced .

ونظراً لأهمية هذان النوعان من التفسيرات فى مختلف مجالات القياس والتقويم التربوى والنفسى وفى تصميم وبناء أدوات القياس وتحليل نتائجها، فسوف ينالا قدراً وافيًا من المناقشة فى فصول لاحقة، وبخاصة فى الفصلين الثامن والخامس عشر.

التقويم Evaluation، والتقييم Assessment :

كثيراً ما يحدث خلط بين مفاهيم القياس والتقويم والتقييم، فالقياس كما رأينا يصف السلوك وصفًا كميًا، ويحدد مقدار سمة معينة أو أكثر لدى الفرد، ولا يتخطى ذلك إلى إصدار حكم معين على السلوك أو السمة المقاسة، فالدرجة ١٣٠ التى حصل عليها الطالب فى المثال السابق تعد «قياسًا» لسمة معينة يقيسها الاختبار، أى تحدد كم السمة، ولكن إذا أصدرنا حكمًا قيميًا على هذه الدرجة، أو اتخذنا قراراً بشأن الطالب الذى حصل على هذه الدرجة استناداً إلى معيار أو محك أداء معين، فإننا بذلك نكون قد تخطينا وصف الدرجة إلى تقويمها والحكم على مستوى الأداء. فإذا كان محك الأداء المقبول ١٢٠ مثلاً، فإن الدرجة ١٣٠ تفوق المحك ونحكم عليها بأنها درجة مقبولة، أما إذا قررنا أن الدرجة ١٣٠ تناظر تقدير أداء جيد مثلاً، فإننا بذلك نكون قد أصدرنا حكمًا كيفيًا على الدرجة أو الأداء الذى تُمثله.

فالقياس Measurement إذن هو الوصف الكمى الموضوعى للأداء Evalua- tion ، فهو الحكم الكيفى الوصفى على الدرجة ممثلاً فى التقدير النوعى للأداء، وهذا الحكم يفيد فى اتخاذ قرار معين بشأن الفرد الذى حصل على الدرجة، أو اقتراح إجراء مناسب له.

وربما يظن البعض أن التقويم بهذا الشكل عملية بسيطة يمكن لأى فرد غير متخصص أو متمرس إجراؤها، ولكن إذا أمعنا النظر نجد أنه عملية منهجية تتطلب جمع بيانات موضوعية ومعلومات صادقة من مصادر متعددة باستخدام أدوات قياس متنوعة فى ضوء أهداف محددة بغرض التوصل إلى تقديرات كمية، وأدلة كيفية يستند إليها فى إصدار أحكام أو اتخاذ قرارات مناسبة تتعلق بالأفراد.

ومما لا شك فيه أن هذه القرارات لا ينعكس أثرها على الفرد فقط، وإنما على مستوى أدائه ودرجة كفاءته في القيام بأعمال أو مهام معينة. كما أن التقويم لا يقتصر على الأفراد، وإنما يمتد بحيث يشمل تقويم المهام والبرامج والمشروعات والمؤسسات للتحقق من أثرها وفاعليتها في تحقيق أهدافها المحددة.

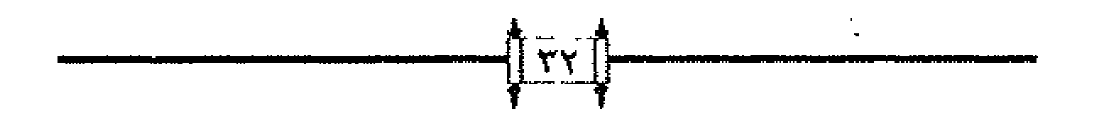
ومن المفاهيم الأخرى شائعة الاستخدام في مجال قياس الشخصية الإنسانية مفهوم «التقييم أو التقدير Assessment». وعلى الرغم من أن هذا المفهوم يستخدم في كثير من الأحيان مرادفًا لمفهوم التقويم، إلا أنه أكثر خصوصية من مفهوم التقويم، وأكثر اتساعًا من مفهوم القياس.

فمفهوم التقييم أو التقدير Assessment استخدمه في مستهل نشأته مجموعة من علماء النفس الذين كانوا يعملون في مكتب الخدمات الاستراتيجية بالولايات المتحدة الأمريكية، حيث أوكل إليهم انتقاء الأفراد الذين يصلحون للقيام بمهام خاصة فيما وراء البحار في الحرب العالمية الثانية. وقد ظهر هذا المفهوم لأول مرة في كتاب «تقييم الإنسان» الذي أصدره هذا المكتب عام ١٩٤٨. وعقب انتهاء هذه الحرب تبنى هؤلاء العلماء المفهوم نفسه في دراساتهم وبحوثهم مما أدى إلى انتشار المفهوم منذ عام ١٩٥٣، وظهوره كمصطلح سيكولوجي في مراجع علم النفس والقياس النفسي.

ويعرفساندبرج(Sundberg , 1977)مفهوم التقييم أو التقدير بأنه مجموعة من العمليات التي تستخدم بواسطة أخصائيين متمرسين؛ للتوصل إلى تصورات وانطباعات واتخاذ قرارات واختبار فروض تتعلق بنمط خصائص فرد معين يحدد سلوكه أو تفاعله مع بيئته. وهذا التقييم يمكن أن يتضمن أساليب متعددة وأدوات متنوعة يعتمد بعضها أحيانًا على القياس الكمي (مثل درجات اختبارات التحصيل، والذكاء، والاستعدادات، والتفكير الابتكاري، ومقاييس الحاجات والميول والاتجاهات والقيم والشخصية)، ويعتمد البعض الآخر عادة على التقديرات الكيفية، والأحكام الوصفية (مثل تقييم استمارات القبول، وتقييم خطابات التوصية، والمقابلة الشخصية)، وذلك بغرض انتقاء الأفراد، وكذلك الملاحظة الكلينيكية، وتقديرات المعلمين للطلاب وبالعكس.

لذلك يقترن مفهوم التقييم أو التقدير Assessment عادة بالمجالات والنظريات الكلينيكية، حيث يميل الكلينيكيون إلى استخدام هذا المفهوم في فحوصهم وتشخيصاتهم لاضطرابات الشخصية التي يعاني منها بعض الأفراد.

وعمومًا، فإن عمليات التقويم والتقييم تعتمد كما ذكرنا على أدوات قياس متنوعة بعضها يؤدي إلى مقادير كمية، والبعض الآخر يعتمد على أحكام الاختصاصيين فيما



يتعلق بسمات الفرد أو شخصيته كفرد متكامل ومتميز عن غيره من الأفراد، وإجراء عمليات التقويم أو عمليات التقييم يعتمد بلا شك على الهدف منها، ونوع القرار المناسب لهذا الهدف وانتقاء أدوات القياس التي تسترشد هذه العمليات بنتائجها. وهذا يتطلب معرفة الاستخدامات المختلفة لأدوات القياس، وأنواعها أو تصنيفاتها.

استخدامات أدوات القياس والتقويم :

تستخدم أدوات القياس التربوى والنفسى فى أغراض متعددة يصعب حصرها نظراً لتنوع الظروف والمواقف التى تتطلب تطبيق هذه الأدوات للاستناد إلى نتائجها فى اتخاذ قرارات تتفق وهذه الظروف والمواقف.

فالباحث التربوى والنفسى يحتاج فى تصميم دراسته وتنفيذها إلى العديد من أدوات القياس المناسبة التى تمكنه من التحقق من فروض دراسته سواء كانت الدراسة تجريبية أو ارتباطية أو وصفية أو كلينيكية، كما يحتاج إليها الممارسون ومتخذو القرارات فى مختلف المجالات التربوية والنفسية مثل المعلم، والاختصاصى النفسى والاجتماعى، والقائم بالعلاج النفسى، والمسئولون عن انتقاء الأفراد لشغل مختلف أنواع الوظائف والمهن أو تصنيفهم وفقًا لمحكات معينة، والمهتمون بتقويم المشروعات والبرامج والمؤسسات وغيرهم.

ولذلك فإن أدوات القياس تخدم الأغراض البحثية والأكاديمية النظرية، كما تخدم الأغراض العملية والتطبيقية.

وسوف نوضح فيما يلى بعض الاستخدامات الرئيسة لأدوات القياس المختلفة.

بناء وتطوير النظريات التربوية والنفسية : فكثير من النظريات فى العلوم السلوكية اعتمدت فى بنائها أو تطويرها على تطبيق أدوات قياس متنوعة، فنظريات التكوين العقلى ونظريات الشخصية فى علم النفس استندت إلى فروض تم التحقق منها بتطبيق مجموعات كبيرة من الاختبارات والمقاييس على عينات مختلفة من الأفراد، وتحليل البيانات المستمدة من هذه الأدوات باستخدام الأساليب الإحصائية، كما أن المفاهيم والتكوينات الافتراضية التى تتضمنها كثير من النظريات التربوية والنفسية يجرى فحصها والتوصل إلى تعريفات إجرائية مناسبة لها عن طريق الاختبارات والمقاييس. وكذلك أسهمت أدوات القياس فى تطور علم القياس التربوى والنفسى Psychometry ذاته عن طريق تصميم وبناء اختبارات ومقاييس جديدة متنوعة، وتطوير المقاييس المتوافرة ، والتوصل إلى صيغ رياضية وإحصائية توضح العلاقات القائمة بين المتغيرات والعوامل المتعلقة بأدوات القياس وخصائصها السيكومترية.

انتقاء الأفراد : تستخدم أدوات القياس في غالبية الأحيان في تزويد المعنيين بشئون الأفراد ببيانات مفيدة يمكن الاستناد إليها بقدر محدد من الثقة في اتخاذ قرارات تتعلق بانتقاء الأفراد Selection لشغل وظائف أو مهن معينه، أو انتقاء الطلاب للالتحاق بالكليات والمعاهد ومراكز التدريب المهني وغيرها، فإلى جانب ما تحدده هذه المؤسسات التربوية من حد أدنى من المتطلبات، فإن الأماكن الشاغرة فيها تقل عن عدد المتقدمين. فهنا يمكن أن تُطبق اختبارات تقيس مدى توافر هذه المتطلبات لدى الطلاب، وتسمى اختبارات مرجعية المحك Criterion - Referenced Tests، كما يمكن تطبيق اختبارات للتنبؤ بنجاح الطلاب أو رسوبهم في الدراسة أو العمل بأقل قدر ممكن من الخطأ. والمقصود بالخطأ هنا قبول طلاب يتبين رسوبهم خلال دراستهم، أو رفض طلاب كان من الممكن نجاحهم لو أتيحت لهم فرصة القبول، وهذه الاختبارات التنبؤية ينبغي أن تقلل احتمال الوقوع في أي من هذين النوعين من الخطأ قدر الإمكان، وتسمى هذه الاختبارات مرجعية الجماعة أو المعيار Norm - Referenced Tests. وهذا يتطلب مراعاة أن يكون هناك تناظر كبير بين المفردات التي تشتمل عليها الاختبارات التي تستخدم في الانتقاء وبين المتطلبات والمهارات والقدرات المطلوبة للنجاح في الدراسة أو العمل، وإجراء العديد من الدراسات التنبؤية قبل استخدام الاختبارات في هذا الشأن.

تسكين الأفراد : إن انتقاء الأفراد للالتحاق بالدراسة بإحدى الكليات او لشغلهم وظيفة أو مهنة معينة يعني المزاوجة بين الفرد والعمل، أي وضع الشخص المناسب في المكان المناسب. ولكن التسكين Placement يعني وضع الطالب الذي تم قبوله بإحدى الكليات مثلا في قسم معين يناسبه بحيث يحقق فيه تفوقًا وتميزا، أو وضع التلميذ الذي تم قبوله في المدرسة في برنامج تعليمي معين يناسبه مثل البرامج العلاجية، أو العادية، أو الإثرائية. وكذلك وضع الفرد الذي تم انتقاؤه في أحد الأقسام المتعلقة بالمهنة أو الوظيفة التي شغلها. ويختلف التسكين عن الانتقاء في أن قرارات الانتقاء تتضمن قبول الفرد أو رفضه، فإذا ما رُفض عليه أن يبحث عن مكان آخر، أما إذا تم قبوله فإن المؤسسة تعمل على تسكينه وفقا لإمكاناته واحتياجاته. فالتلميذ الذي يقبل في مدرسة مهنية معينة ينبغي أن تعمل المدرسة على تسكينه في قسم مناسب لمهنته، أو تُقدِّم له برنامجا تدريبيا علاجيا، أو تقوم بترفيعه في مستوى متقدم بحسب المعلومات التي تقدمها اختبارات التسكين التي تسمى عادة اختبارات تحديد المستوى.

لذلك فإن الاختبار الذى يستخدم فى اتخاذ قرارات الانتقاء ربما لا يصلح فى اتخاذ قرارات التسكين، إذ ربما يصلح اختبار يقيس الاستعداد الدراسى فى التنبؤ بالنجاح فى الكليات، ولكنه ربما لا يصلح فى تسكين الطالب المقبول فى البرنامج التعليمى الأكثر ملاءمة له.

تصنيف الأفراد: على الرغم من أوجه الشبه بين التصنيف والتسكين والانتقاء إلا أن نيتكو (Nitko, 1983) يرى أن التصنيف Classifiation أكثر عمومية منهما. فقرارات التصنيف تتعلق بتعيين الأفراد فى أقسام أو مجموعات مختلفة، فعندما يتم انتقاء الأفراد للالتحاق بالخدمة العسكرية مثلاً يقوم المسئولون بتصنيفهم فى وحدات عسكرية مختلفة وتسكينهم فى أعمال مناسبة لهم داخل هذه الوحدات، كما أن مرضى الاضطرابات العقلية يجرى عادة تصنيفهم، وكذلك الأطفال من ذوى الإعاقات مثل الأصم، والأصم - الأعمى، وصعوبة السمع، والتخلف العقلى، ومتعدد الإعاقات، وشديد الاضطراب الانفعالى، والمعاق بصريًا وهكذا. لذلك يقترح نيتكو Nitko أن التصنيف يتعلق بتعيين الأفراد فى أقسام أو أعمال أو برامج. فإذا كانت هذه الأقسام تمثل مستويات مرتبة لنوع معين من البرامج أو الأعمال (مثل المستويات العلاجية والعادية والإثرائية) فإن هذا التعيين يسمى «تسكين» ، وإذا كان أحد هذه الأقسام يمثل الرفض والآخر يمثل القبول، فإن التعيين يسمى « «انتقاء»، أما «التصنيف» فيمثل أقساما غير مرتبة بطبيعتها ، والقرارات المتعلقة بهما جميعًا تعد من نوع القرارات الإدارية.

التشخيص والعلاج: تستخدم أدوات القياس النفسى والتربوى فى تشخيص الاضطرابات النفسية Diagnosis، وكذلك فى تشخيص جوانب القوة والضعف لدى الطلاب وأصحاب المهن من أجل اقتراح أساليب علاج مناسبة، فمعلم التعليم الأساسى مثلاً يعمل على تشخيص جوانب الضعف لدى تلاميذه فى المهارات الأساسية المتعلقة بالقراءة والكتابة والتخاطب والحساب، والمرشد النفسى يقوم بتشخيص المجالات السلوكية التى يعانى منها الفرد . فالمعلم يحاول معاونة تلاميذه فى تحسين أدائهم، والمرشد النفسى يعمل على تقديم برنامج علاجى مناسب للفرد الذى يعانى من مشكلات نفسية معينة وهكذا. ومن الجدير بالذكر أن الاختبارات التربوية والنفسية التى تستخدم فى اتخاذ قرارات التشخيص تختلف عن الاختبارات التى تستخدم فى الانتقاء والتسكين والتصنيف فى أنها تكون أكثر تفصيلاً وعمقًا نظرًا لتركيزها على جوانب معينة فى السمات المراد قياسها وتعرف جوانب الضعف فيها، ولذلك تسمى الاختبارات التشخيصية Diagnostic Tests ، ويتطلب بناء هذه الاختبارات جهدًا كبيرًا وخبرة فنية عالية فى تقنيات بنائها. ومع أن مؤلف هذا الكتاب قد أفرد مجلدًا خاصًا لهذا النوع من

الاختبارات (علام ، ١٩٩٥) إلا أننا سوف نلقي عليه مزيداً من الضوء في الفصل الثامن، وذلك لأهميته وضرورته في مختلف المجالات التربوية والنفسية وفي كثير من التطبيقات العملية.

الإرشاد والتوجيه التربوي والمهني: فأدوات القياس المتنوعة يمكن استخدامها استخداماً وظيفياً فاعلاً في عمليات الإرشاد Counseling والتوجيه Guidence التربوي والنفسي. وعلى الرغم من أن قرارات الانتقاء والتسكين والتصنيف تعد من القرارات الإدارية المؤسسية، إلا أن قرارات الإرشاد والتوجيه تتعلق بفرد معين شأنها شأن قرارات التشخيص والعلاج. فقد سبق أن ذكرنا أن المؤسسة تتخذ قراراتها في الانتقاء مثلا بما ينطبق على كثير من الأفراد المتقدمين في ظروف متشابهة، بحيث تكون هناك درجة معينة من الثقة فيما يترتب على هذه القرارات، فاحتمال نجاح الأفراد في دراستهم أو في أعمالهم بدرجة كبيرة هو المحك الذي تستند إليه هذه القرارات، أما القرار المتعلق بفرد معين فيؤثر على هذا الفرد بذاته ولا يؤثر فيما عداه، وعليه أن يتحمل المخاطرة في حالة عدم التزامه بهذا القرار.

فالفرد الذي يسعى إلى المرشد التربوي أو النفسي يود أن يتعرف على إمكاناته واستعداداته ويحدد مجال العمل أو الدراسة الأكثر ملاءمة له. لذلك فهو يلجأ إليه بمحض إرادته ويتعاون معه تعاونًا مثمراً وبخاصة في المواقف الاختبارية.

وقرارات التوجيه والإرشاد التربوي والنفسي متعددة، فالتوجيه والإرشاد التربوي يهتم بتعريف الطالب بما يمكنه دراسته من مقررات والتخصص الدراسي الذي يناسبه. والتوجيه والإرشاد المهني يهدف لمعاونة الفرد على اختيار مجال عمل يستطيع أن يؤديه أداء متميزاً بما يتفق واستعداداته وسماته الشخصية، مثل مجال الطب أو الهندسة أو الأعمال الحرة أو الفنية أو الحرفية. والتوجيه والإرشاد الشخصي يركز على المشكلات الشخصية لدى الفرد التي ربما تعكس اضطرابات في الشخصية مثلا أو تكون ناتجة عن سوء تكيف وقتي، مثل سوء تكيف المراهق مع أقرانه الأكثر تفوقـًا، أو سوء تكيفه مع الجنس الآخر.

وفي جميع الحالات ربما لا يقتصر المرشد التربوي أو النفسي على استخدام الاختبارات، وإنما يمكنه أن يجمع معلومات ثرية من مصادر متنوعة. إذ ربما يقوم بمقابلات شخصية تفصيلية مع العميل لتعرف وجهات نظره وخلفيته واتجاهاته ورغباته وأهدافه وخصائصه، وربما يلجأ إلى جمع معلومات عن الفرد من مصادر أخرى ذات علاقة به مثل السجل المدرسي، أو آراء أقرانه ومعلميه ووالديه وأصدقائه ، وغيرذلك.

وانتقاء الأدوات المناسبة يعتمد على خبرة المرشد النفسى بهذه الأدوات وتمرسه على تطبيقها والإفادة من نتائجها لكى يتخذ قراراً صادقًا فيما يتعلق بالعميل.

وربما يتساءل القارئ عن أوجه الاختلاف بين قرارات الانتقاء والتشخيص والإرشاد والتوجية . والحقيقة أن الإرشاد والتوجيه يعدان نوعًا من التشخيص باستخدام سلسلة متتابعة من أدوات القياس المتنوعة التى أشرنا إليها أثناء عملية الإرشاد التى قد تستغرق عدة جلسات.

وعلى الرغم من أن الانتقاء والإرشاد يستندان إلى عمليات أساسية متشابهة حيث يهتم كلاهما بالتنبؤ بالنجاح، إلا أن الانتقاء يركز بدرجة أكبر على عملية التنبؤ، كما يستندان إلى نتائج أدوات قياس متنوعة على الرغم من اختلاف الإطار المرجعى لكل منهما، فقرارات الانتقاء تنتهى بقبول فرد له خصائص معينة فى دراسة أو مهنة معينة مناسبة لخصائصه من بين أعداد كبيرة من المتقدمين للدراسة أو المهنة نفسها بحيث يتوقع أن يكون هذا الفرد أفضل المتقدمين. وعلى العكس فإن قرارات التوجيه والإرشاد تهدف لتحديد مجال العمل المناسب للفرد بحيث يستثمر فيه إمكاناته وقدراته استثماراً فاعلا. وتتأثر قرارات التوجيه والإرشاد بمدى دقة أدوات القياس فى التنبؤ بالأداء فى عمل معين، ولذلك يفضل التنبؤ بالأداء فى مجال عمل محدد بدلاً من التخصيص، حيث إن الفرد ربما يمكنه أداء أكثر من عمل واحد بفاعلية، فالنجاح فى أداء الأعمال يعتمد إلى جانب الاستعدادات على المناخ السائد فى العمل، واتجاهات المشرف على العمل، وسياسة المؤسسة، والعلاقات بين الزملاء، وتوافر الأجهزة المناسبة.

تقويم تحصيل الطلاب:

كثيراً ما يقترن تقويم الطلاب بتقويم التحصيل الدراسى، ولكن التحصيل الدراسى يعد واحداً من المتغيرات الكثيرة التى يتضمنها هذا التقويم، فهناك العديد من المتغيرات التى ينبغى أن يأخذها المعلم بعين الاعتبار عند تقويمه لطلابه. فالتحصيل الدراسى يرتبط بمتغيرات عقلية أو معرفية أخرى،مثل الذكاء والاستعدادات بأنواعها والأنماط المعرفية، كما يتضمن المتغيرات الوجدانية مثل الميول والاتجاهات والقيم والسمات المزاجية، والمتغيرات الحركية المهارية.

لذلك فإن أدوات تقويم الطلاب لا تقتصر على الاختبارات التحصيلية بنوعيها الرئيسين، أى الاختبارات مرجعية الجماعة أو المعيار والاختبارات مرجعية المحك، وإنما ينبغى أن تتسع هذه الأدوات بحيث تشمل اختبارات الذكاء والاستعدادات ومقاييس الميول والاتجاهات وغيرها من الأدوات متضمنة بطاقات ملاحظة السلوك، وموازين التقدير، والمقابلات الشخصية، ومؤشرات الأداء، وقوائم المراجعة، واختبارات المواقف.

والبيانات المستمدة من هذه الأدوات المتنوعة لا تفيد فقط فى تقويم ومتابعة التقدم الدراسي للطلاب وإنما تفيد أيضا فى تقويم عملية التعلم، وبعبارة أخرى فإن عملية تقويم الطلاب تقدم تغذية مرتدة للطالب والمعلم تساعد فى زيادة دافعية الطالب وتشجيعه على تكوين عادات استذكار جيدة، وتعريفه بجوانب القوة والضعف فى تحصيله أو أدائه. وبذلك تؤدى عملية تقويم الطلاب إلى تحسين وإثراء عملية التعليم، حيث إن هاتين العمليتين متكاملتان، وهذا يؤكد أهمية التقويم البنائى الذى سوف نوضحه بعد قليل.

أدوار التقويم وأهدافه:

إن التقويم بمفهومه العلمي ـ كما أوضحنا ـ يختلف بالطبع عن التقويم الذى يمارسه الإنسان العادى فى حياته اليومية. فالإنسان العادى يقوِّم العديد من الأشياء والأمور التى يواجهها فى حياته اليومية، فهو يقوِّم الأطعمة التى يأكلها،والأشياء التى يشتريها،والكتب التى يقرأها، والأماكن التى يزورها وغير ذلك. ويتضمن هذا التقويم نوعًا من الحكم على الشيء المطلوب على الرغم من أنه ليس تقويما علميًا بالمعنى المتعارف عليه.

ونحن فى المجال التربوى مثلاً نقابل كثيراً من المواقف التى نمارس فيها هذا النوع من التقويم الذى يمكن أن نطلق عليه التقويم غير الرسمى، وهو يتراوح بين حكم المعلم على حسن تهوية الفصل وإضاءته إلى تقدير مدير المدرسة لفاعلية وكفاءة أعضاء الهيئة التدريسية.

ولكن يصعب الركون إلى نتائج هذا النوع من التقويم فى اتخاذ القرارات التربوية المختلفة، ومن هنا كان لابد من الاعتماد على التقويم العلمى المنظم بمفهومه الذى قدمناه.

ومفهوم التقويم من هذا المنظور يمكن أن يُطبق فى مختلف المواقف التربوية سواء كان تقويم أفراد أو تقويم مشروعات وبرامج، فمثلا يمكن استرشادًا بهذا المفهوم تعرف مدى فاعلية أنشطة تعليمية معينة وفقًا لمحكات محددة، ومعرفة ما إذا كان تأثير نشاط أو برنامج معين أفضل من تأثير نشاط أو برنامج آخر، وكذلك تعرف المتغيرات التى تأثرت من جراء تنفيذ نشاط أو برنامج معين على جماعة ذات خصائص معينة، وما إذا كان استخدام أو تنفيذ برنامج معين يوازى كلفة إعداده وهكذا.

وبذلك تتعدد أدوار التقويم في المجالات التربوية، فللتقويم دور في إعداد المعلمين وأنشطتهم، وفي عملية بناء المناهج الدراسية، وفي التجارب الميدانية التجديدية المتعلقة بتحسين عمليات التعليم والتعلم، وفي انتقاء أفضل الاستراتيجيات والتقنيات التربوية، وفي الحكم على استمرار برنامج تربوي معين أو تعديله أو إلغائه. كما أن للتقويم دور بالغ الأهمية في متابعة التقدم الدارسي للطلاب وتحديد مستواهم التحصيلي وتوجيههم التوجيه التعليمي والمهني. ولذلك لا ينبغي أن نخلط بين أهداف التقويم وأدواره، فعادة يكون للتقويم أدوار متعددة بل ويجب أن يكون التقويم هكذا بالنسبة لبرنامج تعليمي معين، ولكن ربما يرتبط بكل دور من هذه الأدوار العديد من الأهداف. فمثلاً في عملية بناء منهج معين، ربما يكون أحد أدوار التقويم تحسين وتطوير المنهج أثناء عملية بنائه، ويرتبط بهذا الدور العديد من الأهداف أو الأسئلة التي تتطلب إجابة مثل: هل أهداف المنهج واضحة وضوحًا تاما في أذهان القائمين ببنائه؟ وهل هم متحمسون لبنائه؟ وهل تسير عمليات البناء وفق هذه الأهداف بالفعل؟ وهكذا. وربما يكون لعملية التقويم في هذا الصدد دور آخر يتعلق بمساعدة المسئولين عن تنفيذ المنهج في اتخاذ قرار بشأن تبرير استخدامه في مرحلة تعليمية معينة وفقًا للنتائج التي أبرزتها عملية التقويم في الدور السابق.

أي أن دور التقويم يتعلق بالنشاط المراد تقويمه، أما أهداف التقويم فتتعلق بالأسئلة المراد الإجابة عنها والتي تتبلور حول إمكانية وفاعلية إعداد وتصميم وتنفيذ هذا النشاط.

ولعل هذا التمييز بين أدوار التقويم وأهدافه يزيل التخوف والقلق الذي يعتري كل من يسمع كلمة «تقويم Evaluation» ، ويؤكد الجانب الإيجابي البنّاء لهذا المفهوم، فكثيرًا ما ارتبط التقويم بالكشف عن مواطن الضعف والقصور سواء في الأدوات أو الأفراد أو في الإمكانات المادية والبشرية.

لذلك فإن للتقويم دورين رئيسين هما الدور البنائي التشخيصي Formative Evaluation والدور الختامي التجمعي Summative Evaluation. فالتقويم البنائي التشخيصي يهدف للكشف عن جوانب القوة والضعف في برنامج تعليمي معين أثناء إعداده وتنفيذه في مرحلة التجريب بغرض مراجعة مكونات البرنامج وتعديلها وتطويرها، وكذلك تشخيص مواطن القوة والضعف في أداء المتعلمين وتحديد الصعوبات التي يواجهها كل منهم أثناء التعليم واتخاذ ما يلزم من أساليب العلاج. ومساعدة المتعلم في التعرف على قدراته وإمكاناته، واقتراح سبل ووسائل تحسينها وتنميتها إلى أقصى حد ممكن.

أما دور التقويم الختامي التجمعي فهو الكشف عن مدى تحقيق البرنامج لأهدافه بعد إتمام إجراءات التقويم البنائي المستمر، وكذلك تحديد المستوى التحصيلي العام للمتعلمين في نهاية مدة دراسية معينة بالنسبة للمعارف والمهارات والاتجاهات المتعلقة بالمجالات الدراسية؛ وذلك لاتخاذ قرارات تتعلق بتقديراتهم ومنحهم الشهادات الدراسية، وهذا يؤكد أيضًا الدور الإيجابي لعملية التقويم الختامي بنفس القدر الذي يؤكد دور عملية التقويم البنائي.

ولتوضيح ذلك نفترض أنه تبين من نتائج تطبيق هذه الأدوات المتعددة أن عدداً قليلاً من الطلاب لم يحقق هدفًا تعليميا معينا أو مجموعة من الأهداف المتعلقة بمحتوى دراسي معين، فإنه ربما يستنتج المعلم أن عملية التعليم كانت فاعلة، أما إذا تبين أن كثيراً منهم أخفق في تحقيق مجموعة كبيرة من الأهداف التعليمية، فإنه ربما يقرر أن عملية التعليم التي اتبعت كانت قاصرة. وفي كلتا الحالتين ينبغي أن يتخذ المعلم قراراً معينًا فيما يتعلق بأسلوب التعليم الذي اتبعه. كما أن التقويم البنائي Formative Evaluation للطلاب يقدم تغذية مرتدة فيما يتعلق بتخطيط المعلم المستقبلي للأنشطة التعليمية. فربما يجد المعلم أن معدل سرعته في عرض المادة يجب أن يقل أو يزداد، أو ربما يكشف التقويم عن حاجة المتعلمين إلى تعلم بعض المهارات اللازمة المتابعة موضوع مستقبلي معين، ويطلق على هذه المهارات «المهارات المدخلية Entry Skills» أو المتطلبات اللازمة للبدء في دراسة معينة Prerequisites، وقياس الأهداف المدخلية اللازمة لتعلم الأهداف المرتبطة بها، وتكون عادة من نوع الاختبارات مرجعية المحك Criterion - Referenced Tests ، كما يحتاج المعلم أيضًا إلى اتخاذ قرارات تتعلق بتقديرات الطلاب في نهاية مدة دراسية معينة لتحديد المستوى التحصيلي العام لكل منهم، وهنا تبرز أهمية التقويم الختامي Summative Evaluation ، وبالطبع تعتمد هذه التقديرات على نتائج التقويم المدخلي والتقويم البنائي المستمر أثناء تنفيذ عملية التعليم. ويجب أيضًا أن تُستخدم أدوات التقويم الختامي التي تكون عادة من نوع الاختبارات مرجعية الجماعة أو المعيار Norm - Referenced Tests إذا كان الهدف من هذه التقديرات توجيه الطلاب إلى مرحلة تعليمية تالية أو دراسة مستقبلية معينة. ويعرف هذا النوع من الاختبارات بالاختبارات التحصيلية المقننة Standardized Achievement Tests التي تحدد المركز النسبي للطالب في مجموعته العمرية أو فرقته الدراسية فيما يتعلق بمستواه التحصيلي العام.

وفى أى الأحوال يجب أن تزود أدوات القياس والتقويم كل من المعلم والطالب ببيانات صادقة ومتسقة عن تحصيل الطلاب بمفهومه الشامل لما لهذه البيانات من أهمية فى اتخاذ القرارات التعليمية المختلفة التى تمس حاضرهم ومستقبلهم.

وهذا يوجه النظر إلى أهمية تصميم وبناء الاختبارات التحصيلية التشخيصية واختبارات الكفايات، والاختبارات المقننة في مختلف مجالات الدراسة.

تقويم المناهج والبرامج والمشروعات التربوية المستحدثة:

فالاختبارات والمقاييس يمكن أن تستخدم استخدامًا فاعلاً فى اتخاذ قرارات تتعلق بالمناهج والبرامج والمشروعات التجديدية التى تعمل المؤسسات التربوية على تصميمها وتنفيذها. وقد اختلف العلماء فى كيفية تصميم المنهج الدراسى Curriculum وطرق بنائه، فمنهم من اعتبر أن عملية تصميم المنهج تشتمل على: تشخيص الحاجات التربوية وتقييمها، وصياغة الأهداف التعليمية العامة والأهداف السلوكية، وانتقاء المحتوى، وإعداد الخبرات التعليمية وتنظيمها، وتحديد طرق وأساليب تقويم فاعلية ما يتم تعلمه، وهذه العمليات تقدم تحديدًا وتعريفًا عريضًا وشاملاً لعمليات تصميم المنهج الدراسى وبنائه وتقويمه.

وفى ضوء هذا التحديد تصبح عملية تقويم المنهج هى عملية تقويم للبرنامج التعليمى برمته، بما يشتمل عليه من أساليب واستراتيجيات تعليم، وكتب مدرسية، وتقنيات تربوية. وهذا التقويم لا يجب أن يتم بطريقة مجردة، وإنما فى مواقف مدرسية فعلية بحيث يصبح تقويم المنهج عملية ديناميكية مبدعة. وهذا يتطلب وجود محكات داخلية وخارجية للحكم على المنهج. فالمحكات الداخلية تفيد في معرفة ما إذا كان المنهج المعين قد حقق أهدافه المرجوة، والمحكات الخارجية تفيد فى معرفة ما إذا كان المنهج المعين أفضل من المناهج الأخرى المناظرة له من حيث تأثيره واتجاه المعلمين والطلاب نحوه ومردود كلفته وفاعليته.

وهذا النوع من التقويم الذى يعتمد على محكات خارجية يصعب إجراؤه نظرًا لصعوبة إمكانية تطابق منهجين فى الأهداف حتى لو ارتبط كل منهما بالمحتوى الدراسى نفسه، كما يصعب وجود أدوات قياس تكون صادقة بالدرجة نفسها لقياس فاعلية كل من المنهجين؛ لذلك يجب البدء بفحص أهداف كل من المنهجين فحصًا دقيقًا إذا أردنا الموازنة بينهما، وبناء أدوات قياس صادقة بالنسبة لكل منهما إذا لم تكن متوافرة.

وفيما يتعلق بتقويم البرامج والمشروعات التربوية المستحدثة، فإن التقويم يلعب دورًا مهمًا في كل مرحلة من مراحل تصميم وإعداد وتنفيذ هذه البرامج. ونقصد بالبرنامج أو المشروع التربوى مجموعة من الأنشطة أو العمليات المستحدثة الهادفة والمقصودة التى يمكن تحديدها بدقة ويفترض أنها تؤثر تأثيرا معينًا فى مجموعة أو مجموعات مستهدفة من الأفراد. ومن أمثلة ذلك برنامج مستحدث لتعليم الأطفال القراءة باستخدام التقنيات السمعية والبصرية، وبرنامج تجديدى لتعليم المواطنة المنتجة الواعية عن طريق التدريبات الحياتية، أو برنامج تربوى مبتكر لتعزيز القيم العالمية والتسامح والديمقراطية داخل المدرسة، أو برنامج مستحدث لإعداد مواد تعليمية ملائمة للمرحلة الثانوية فى مجال التسامح والتفاهم الدولى، وغير ذلك من البرامج التربوية التجديدية التى تعد الأفراد للحياة المثمرة فى القرن المقبل.

فالتقويم بما يتضمنه من عمليات كشف وجمع للبيانات والمعلومات باستخدام أدوات قياس متنوعة يساعد القائمين على تخطيط هذه البرامج وتنفيذها وإدارتها فى توجيه مسار هذه البرامج وتطوير مكوناتها عن طريق التقويم البنائى Formative Evaluation لتحقيق أهدافها المرجوة، إذ ربما تكشف عملية التقويم البنائى عن قصور فى نظام تمويل البرنامج أو نظام إعداد المشرفين عليه، أو نظام المواد والأدوات، أو المدة الزمنية اللازمة لتنفيذ كل مرحلة من مراحل البرنامج، أو غير ذلك مما يؤثر فى فاعلية البرنامج وفى تحقيقه لأهدافه. فالتقويم البنائى يفيد فى اتخاذ قرارات أفضل فى كل من هذه المراحل. أما التقويم الختامى Summative Evalua-tion فيفيد فى تحديد فاعلية البرنامج ككل وتأثيره بعد أن يكون قد خضع لعمليات التقويم البنائى المستمر أثناء إعداده وتنفيذه استنادًا إلى أدوات قياس متنوعة ومناسبة.

تقويم المعلم: إن فاعلية العملية التربوية تعتمد اعتمادًا أساسيًا على ما حققه المعلم فى الفصل المدرسى، فجميع المصادر التى يستخدمها المعلم سواء كانت مواد أو أدوات أو تقنيات أو غير ذلك يجب أن تفيد فى النهاية فى إحداث تغيرات فى سلوك الطلاب. وقد أكدت حركة المسئولية التربوية Accountability التى انتشرت فى بعض دول العالم المتطور أهمية جمع بيانات باستخدام أدوات قياس متنوعة عن طبيعة هذه التغيرات لتقييم فاعلية المعلم فى ضوئها، كما أكدته التوجهات المعاصرة المتعلقة بتقويم التقدم التربوى على المستويات الدولية والوطنية Educational Progress، إذ يرى رواد هذه التوجهات أن المعلم مسئول عن إحداث تغيرات سلوكية محددة مسبقًا لدى الطلاب، وعليه أن ينظم الخبرات التعليمية لهم بأساليب تيسر اكتسابهم للمعارف والمهارات والاتجاهات المتعلقة بالمجالات الدراسية وبمختلف الأنشطة المدرسية.

لذلك من الأهمية بمكان تقويم عمل المعلم تقويمًا بنائيا للتأكد من فاعلية تنظيمه للبيئة التعليمية بما يحقق التغيرات المرجوة في سلوك الطلاب، وبالطبع يكون المعلم أقدر من غيره على القيام بمثل هذا التقويم. إذ يجب أن يعتمد على أساليب التقويم الذاتي Self Evaluation للتحقق من ذلك، كما يمكن أن يعتمد على ملاحظات الطلاب فيما يتعلق بالتفاعل القائم داخل الفصل المدرسي. ولجعل البيانات الناتجة عن هذه الملاحظات أكثر موضوعية يمكن تصميم بطاقات خاصة للتقويم تدوّن فيها عبارات مرتبطة بمختلف الأنشطة التي ينبغي أن يوفرها المعلم، وكذلك بطبيعة التفاعل الذي يتم بين المعلم وطلابه وغير ذلك من الأنشطة.

ويجب أن تكون هناك محكات واضحة لفاعلية العملية التعليمية يتم في ضوئها تقويم مدى قيام المعلم بأدواره المتغيرة وفاعلية أدائه . فمن المعلوم أن هناك العديد من المتغيرات خارج الفصل المدرسي تؤثر في النواتج التي يهدف المعلم لتحقيقها ولا يكون له سيطرة عليها. كما أن هناك متغيرات ترتبط بفاعلية المعلم مثل الكفايات التي يمتلكها Competencies وتمكنه من مجاله الدراسي وتحمسه وإدارته للفصل وقدرته على التنظيم والإبداع وغير ذلك. لذلك ينبغي مراعاة هذه العوامل إذا أردنا أن يكون نظام تقويم أداء المعلم أكثر واقعية. وهنا يقع العبء غالبًا على المعلم في إعداد أساليب متطورة لتقويم تحصيل الطلاب وإنجازاتهم باستخدام أدوات قياس مناسبة.

تقويم المدرسة: إن المدرسة تعد مؤسسة معقدة التركيب، فهي تشتمل على العديد من المكونات المتشابكة والمتفاعلة التي ينعكس أثرها على المخرجات Outputs المتمثلة في تحصيل الطلاب ومكتسباتهم وإنجازاتهم في مجالات متنوعة ومختلفة تتعلق بالمنهج الدراسي بمفهومه الشامل، وإسهامات المدرسة في البيئة المحلية. وتتأثر هذه المخرجات تأثرًا واضحًا بفريق العمل من معلمين وإداريين وعاملين، وبالمؤسسات الأخرى التي تشارك المدرسة مهامها، والمنهج الدراسي وأهدافه، والمباني التعليمية وغيرها من المصادر. ويتطلب تقويم المدرسة تقويم جميع هذه المكونات التي تعد أيضًا مكونات مركبة. فتقويم المدرسة إذن يتضمن تقويمًا للبرنامج التربوي برمته عن طريق جمع بيانات ومعلومات تتعلق بمختلف جوانب البرنامج ووظائفه.

والهدف من ذلك هو تحديد مدى تحقيق المدرسة لأهدافها المرجوة، وتعرف مجالات القوة والضعف في البرنامج التربوي بعامة. وهذه المعلومات تقدم تغذية مرتدة تساعد المسئولين عن إدارة المدرسة في تحديد الخطط والأنشطة المستقبلية واتخاذ القرارات المتعلقة بتمويل هذه الأنشطة بما يتناسب وأهميتها.

ومن المكونات الأساسية التي تهتم بها كثير من الدول فى تقويم المدرسة ما يسمى بالبرامج الاختبارية School Testing Programs التى تكون عادة برامج متكاملة وشاملة لمختلف المجالات. فعلى الرغم من أن تقويم التحصيل الدراسى للطلاب يعد من مجالات التقويم المهمة للمدرسة ،حيث إن الهدف الرئيسى للمدرسة هو إثراء تعلم الطلاب، إلا أن هناك مجالات أخرى لا تقل عنه أهمية ينبغى تقويمها كما ذكرنا لارتباطها ارتباطًا وثيقًا بتعلم الطلاب. وتقويم المنهج فى ذاته ربما لا يقدم دليلاً كافيًا على أنماط تحصيل الطلاب. ولذلك ينبغى تطبيق مجموعات من الاختبارات والمقاييس التى تقدم صورة شاملة عن الطالب من حيث سمات شخصيته واستعداداته وميوله ومفهومه عن ذاته، حيث إن هذه جميعًا تؤثر تأثيراً بالغًا فى مستوى تحصيله وفاعلية تعلمه. وهذه المجموعات من الاختبارات ينبغى العناية بانتقائها أو إعدادها وتطبيقها وتفسير نتائجها فى ضوء معايير Norms مناسبة؛ وذلك للإفادة من نتائجها فى مراجعة عمل المدرسة. كما ينبغى الإفادة أيضًا من تقويم كفايات المعلمين والإداريين، والموارد المادية والمالية للمدرسة، والمبانى المدرسية فى اتخاذ قرارات تتعلق بالمنظومة التعليمية وأساليب تفعيلها.

تصنيف الاختبارات والمقاييس التربوية والنفسية :

نظرًا لتعقد وتشابك الظواهر التربوية والنفسية وتعدد التطبيقات الميدانية للقياس فى مجالات انتقاء الأفراد لمختلف الوظائف والأعمال والمهام وتسكينهم وتصنيفهم، وفى التوجيه والإرشاد النفسى ، وفى تقويم الأفراد والبرامج، والمؤسسات وغير ذلك، كما سبق أن رأينا، فقد تعددت وتنوعت أدوات القياس والتقويم بحيث تناسب طبيعة الظاهرة المراد قياسها وما تشتمل عليه من متغيرات. لذلك أصبح فى متناول الباحثين والممارسين العديد من أدوات القياس التى تمكنهم أن ينتقوا من بينها أكثرها ملاءمة للأفراد الذين ستطبق عليهم، ولطبيعة السمة المعينة. فأدوات قياس الذكاء والاستعدادات والتحصيل مثلا تختلف عن أدوات قياس القلق والاكتئاب، أو قياس الميول والاتجاهات،أو الجوانب النفسية الحركية. ومما هو جدير بالذكر أنه لا يوجد تصنيف وحيد لهذه الأدوات، وإنما يمكن تصنيفها وفقًا للعديد من الأبعاد. وفيما يلى أحد التصنيفات الممكنة.

أدوات قياس الجوانب المعرفية فى مقابل أدوات قياس الجوانب الوجدانية:
فالجوانب المعرفية للشخصية المتمثلة فى القدرات العقلية والتحصيل تتطلب أدوات قياس يحاول فيها الفرد المختبر بذل أقصى جهده ليحصل على درجة تعبر تعبيراً صادقًا

عن قدرته الفعلية. فالاختبار الذى يقيس الاستعداد اللغوى مثلاً يتطلب من الفرد أن يجيب عن أكبر عدد من المفردات التى تكون عادة من نوع الاختيار من متعدد فى زمن محدد، أما الجوانب الوجدانية والسمات المزاجية للشخصية فإنها تتطلب أدوات لقياس سلوك الفرد الذى يميزه عن غيره من الأفراد، أى سلوكه المعتاد فى المواقف الانفعالية مثلاً. وهذه المقاييس تشتمل على عبارات ليس لها استجابة نمطية محددة وإنما يستجيب الفرد بما يتفق مع خصائصه وسلوكه اليومى المعتاد فى مواقف مشابهة للموقف الذى تشير إليه العبارة. ومن أمثلة ذلك: أحب تنسيق الزهور، أو أحبذ الاختلاط فى الجامعة، أو أعتقد أن الناس لا تحبنى، أو تنتابنى وساوس كثيرة، وهكذا. ونظراً لأن هذه العبارات أو التساؤلات ليس لها إجابة صحيحة أو إجابة خطأ، وإنما تختلف الإجابة باختلاف وجهة نظر أو مشاعر أو معتقدات الفرد، فإن قياس هذه الجوانب يكون أكثر صعوبة من قياس الجوانب المعرفية، حيث يتطلب أدوات قياس تعتمد على أساليب الملاحظة الموضوعية، والتقارير الذاتية، والاستبيانات وغيرها من أساليب التقييم الشخصى Personal Assessment.

أدوات قياس جماعية فى مقابل أدوات قياس فردية: فأدوات القياس الجماعية تطبق على مجموعة كبيرة من الأفراد فى وقت واحد، بينما أدوات القياس الفردية تصمم بحيث تطبق على فرد واحد فى وقت واحد، فاختبارات الذكاء الجماعية يمكن أن تطبق على تلاميذ فصل مدرسى فى آن واحد، أو على المتقدمين للالتحاق بالخدمة العسكرية، وكذلك اختبارات التحصيل واختبارات الالتحاق بالكليات وغيرها. غير أن هناك مواقف تتطلب استخدام أدوات قياس فردية مثل: مواقف الإرشاد والتوجيه التربوى والمهنى أو قياس ذكاء الأطفال، وتطبق هذه الأدوات بواسطة أخصائى متمرس ومدرب تدريبًا جيداً على استخدامها، حيث يقوم بملاحظة توجه الفرد المستجيب وأداؤه، وكذلك يستطيع أن يتحاور معه حول استجاباته لكى يتفهمها الأخصائى جيداً، أى أن أدوات القياس الفردية تسمح بقدر كبير من التفاعل بين الفاحص والمفحوص، وتكوين ألفة بينهما، وبذلك تقدم فرصًا ثرية للملاحظات الكلينيكية. وهناك اختبارات فردية متعددة لقياس الذكاء والشخصية سوف نتناولها فى البابين الثانى والثالث.

ومما لا شك فيه أن أدوات القياس الفردية يستغرق تطبيقها وقتًا طويلاً، وتتطلب غرفة خاصة مجهزة تجهيزاً مناسبًا لتطبيق الاختبار فى ظروف ملائمة؛ لذلك فإن هذه الاختبارات الفردية لا تستخدم إلا إذا كان هناك ضرورة لاتخاذ قرار مهم يتعلق بفرد ما، كأن يترتب على هذا القرار نقل طفل من الفصول المعتادة إلى فصول المتخلفين عقليًا مثلاً، أو تشخيص الاضطرابات النفسية.

أدوات قياس بمثابة عينة فى مقابل أدوات قياس بمثابة مؤشر . تعتمد بعض أدوات القياس وبخاصة اختبارات التحصيل فى بنائها على الانتقاء العشوائى لعينة من المفردات التى تقيس هدفًا أو أهدافًا تعليمية معينة بحيث تمثل هذه المفردات جميع المفردات التى يمكن أن تقيس هذا الهدف أو هذه الأهداف، ومن ثم يمكن الاستدلال على تحقق الهدف أو الأهداف من هذه العينة من المفردات الاختبارية، فمثلا إذا أراد معلم أن يقيس مهارة الطالب فى تعرُّف معنى الرموز الكيميائية، فإنه ينتقى عشوائيًا عينة ممثلة لهذه الرموز ويختبر الطالب بها، ومن هذه العينة الصغيرة يستطيع المعلم الاستدلال على تحقق هذا الهدف لدى الطالب، أى أن الاختبار يقيس عينة سلوكية Sample Test، أما إذا استخدم اختبارًا معينًا للتمييز بين مجموعتين من الأفراد في سمة معينة تختلف عما يقيسه الاختبار بالفعل فإن الاختبار يكون بمثابة مؤشر Sign للتمييز بين المجموعتين فى هذه السمة .

فإذا اشتمل اختبار عقلى معين على مفردات تقيس قدرة التلميذ على تذكر مجموعة من الأرقام في تسلسل معين، وتوصلت الدراسات إلى أن عدم قدرة الفرد على ذلك تميز بين المتخلفين عقليًا والعاديين، ففى هذه الحالة لا يصبح قياس تذكر الأرقام هو الهدف من تطبيق الاختبار، وإنما يستفاد من الاختبار كمؤشر لأغراض معينة.

أدوات قياس لفظية فى مقابل أدوات قياس الأداء: فمعظم أدوات القياس تتطلب استخدام اللغة سواء فى تعليمات المقياس أو مفرداته ، لذلك تسمى أدوات قياس لفظية Verbal Tests مثل معظم اختبارات الجوانب المعرفية والوجدانية. ولكن هناك أنواع من أدوات القياس تستخدم مع الأطفال الصغار والأميين والمعاقين لغويًا، تشتمل على أشكال أو رسوم أو صور أو متاهات أو مواد عيانية، فهذه تسمى أدوات قياس غير لفظية Non - Verbal، إلا أن الفاحص يقرأها جهريًا. ولا تتطلب عادة أدوات قياس الأداء إجابات تحريرية لأسئلة معينة، وإنما تتطلب أداء مطلب يتعلق بالمثيرات العيانية المعطاة، ومثال ذلك بعض الاختبارات الفردية لقياس الذكاء والاستعدادات المكانية والميكانيكية، ومقاييس الجوانب النفسية الحركية.

ومن الجدير بالذكر أن كثيرًا من الاختبارات الفردية للذكاء تتضمن أجزاء لفظية وأجزاء أدائية، وبذلك يمكن قياس الجانب غير اللفظى فى الذكاء ، وكذلك المقارنة بين درجات كل من الجزئين لتعرف جوانب القصور اللغوى لدى الفرد.

أدوات قياس تتطلب السرعة فى مقابل أدوات قياس تتطلب القوة: تتباين أدوات القياس فى درجة تحديدها لزمن الإجابة. فهناك اختبارات موقوتة ، أى يتحدد فيها زمن الإجابة تحديداً دقيقًا ولا يسمح للفرد تخطى هذا الزمن، وعادة يشتمل الاختبار على مفردات سهلة نسبيًا، ولكن نظراً لمحدودية الوقت فإن كثيراً من الأفراد لا يستطيعون إجابة جميع مفردات الاختبار فى الزمن المحدد باستثناء النابهين.

ونظرًا لأن الهدف من هذه الاختبارات قياس سرعة الأداء، فإن تباين درجات الأفراد لا ترجع إلى صعوبة مفرداته، وإنما إلى اختلاف معدل سرعتهم فى الإجابة، وكثير من اختبارات الذكاء والاستعدادات تعد من نوع اختبارات السرعة Speed Tests. أما أدوات القياس التى تهتم بما يستطيع الفرد أداءه بغض النظر عن سرعته، فإنها تسمى اختبارات القوة Power Tests، وهذه الاختبارات تصمم لقياس قدرة الفرد على حل مشكلات تتميز بالصعوبة بغض النظر عن الزمن المطلوب للإجابة. ولعل قياس تحصيل الطلاب يعتمد على هذا النوع من الاختبارات، حيث يكون التركيز على تقييم ما حققه الطالب من معارف ومهارات ، لذلك يعطى فرصة كافية للتفكير والإجابة عن جميع المفردات قدر إمكانه.

والحقيقة أنه لا توجد حدود قاطعة بين نوعى الاختبارات، إذ يمكن أن يكون الاختبار سرعة لمجموعة معينة، وقوة لمجموعة أخرى. فمثلاً اختبار فهم معانى الكلمات يعد اختبار سرعة لطلاب المرحلة الثانوية إذا كانت مفرداته سهلة ومناسبة لتلاميذ المرحلة الابتدائية، ولكنه يعد بمثابة اختبار قوة لهؤلاء التلاميذ.

لذلك ينبغى مراعاة التحديد الدقيق لمثل هذه الاختبارات التى تعتمد جزئيًا على السرعة عند إعدادها.

أدوات قياس موضوعية فى مقابل أدوات قياس ذاتية: فالموضوعية Objectivity والذاتية Subjectivity تتعلقان بطريقة تصحيح الاختبارات والمقاييس التربوية والنفسية وليس بنوع الاختبار أو المقياس. فقد سبق أن أوضحنا مفهوم الموضوعية كأحد مكونات تعريف الاختبار، وذكرنا أنه يعنى عدم تأثر تصحيح الاختبار بالأحكام الذاتية للقائمين بها. فالاختبارات التى تشتمل على مفردات اختيار من متعدد مثلا يتسم تصحيحها بالموضوعية نظرًا لوجود مفتاح تصحيح يطبق على أوراق الإجابات، بل ويمكن تصحيحها بواسطة الحاسوب؛ لذلك لا تختلف درجة الفرد باختلاف القائمين بعملية التصحيح، ولكن يمكن أن تختلف الدرجة التى يحصل عليها الفرد فى اختبار يشتمل على أسئلة مقالية أو أى نوع آخر من الأسئلة التى تتطلب إجابة مفتوحة باختلاف المحكمين أو القائمين بعملية التقدير ،حيث إن عنصر الذاتية يؤدى إلى هذا الاختلاف. ويمكن تلافى ذلك بالعناية بتحديد المطلب الذى يصوغه السؤال

تحديدا دقيقًا لا يشوبه الغموض، وكذلك الاتفاق على محكات أو عناصر أساسية ينبغى توافرها فى الإجابة للاسترشاد بها فى تقدير الدرجات مما يقلل من الذاتية، ويزيد من اتساق الدرجات، وبذلك يكون التصحيح أكثر موضوعية. وعموما، فإن تقنين عمليات تطبيق الاختبار أو المقياس وتعليماته وطريقة تصحيحه يمكن أن تسهم فى تحقيق الموضوعية المرجوة بغض النظر عن نوع أداة القياس المستخدمة.

أدوات قياس يُعِدها المعلم فى مقابل أدوات قياس مقننة: فالمعلم يهتم بقياس مدى تحقيق تلاميذه للأهداف التعليمية للمجالات الدراسية ليس فقط لوضع تقديرات لكل منهم فى نهاية مدة دراسية معينة، وإنما لتبين جوانب القوة والضعف فى تحصيلهم لكى يتسنى له اقتراح أساليب علاج مناسبة للطلاب الذين يجدون صعوبة فى تحقيق هذه الأهداف.

لذلك فإنه يقوم بإعداد اختبارات يكون لها وظيفة تعليمية مهمة فى تحسين دافعية الطلاب وحثهم على بذل المزيد من الجهد لإثراء تعلمهم، وعادة تكون هذه الاختبارات من نوع الاختبارات مرجعية المحك، غير أن هذا يعتمد على تمكن المعلم من المهارات الأساسية فى بناء مثل هذه الاختبارات الصفية.

ولكن المعلم يهتم فى بعض الأحيان بمقارنة تحصيل طلاب فصل أو فرقة معينة في مجال أو أكثر من المجالات الدراسية بمجموعة خارجية ممثلة لمجتمع هؤلاء الطلاب ولتكن مجموعة طلاب الفرقة نفسها على المستوى الوطنى التى تسمى الجماعة المرجعية Norm Group ، فإنه عندئذ ينبغى أن يعتمد على اختبارات تسمى الاختبارات المقننة Standardized Tests . ويقوم ببناء هذه الاختبارات خبراء فى القياس التربوى والمناهج والمعلمين، وتُجرى عملية التقنين عن طريق تطبيق الاختبار وتصحيحه فى ظروف موحدة ومنظمة على الجماعة المرجعية، والتوصل إلى معايير Norms يمكن فى ضوئها مقارنة نتائج مختلف فصول فرقة معينة أو مدرسة معينة بأقرانهم على المستوى الوطنى. فالدرجة التى يحصل عليها أحد طلاب فرقة معينة فى إحدى المدارس فى مجال دراسى معين يمكن مقارنتها بمعيار الجماعة المرجعية الذى تم التوصل إليه لمعرفة المستوى التحصيلى لمجتمع طلاب هذه الفرقة فى المجال نفسه، لذلك فإن مثل هذا الاختبار يكون من نوع الاختبارات مرجعية الجماعة أو المعيار.

ومن الجدير بالذكر أن معظم الاختبارات والمقاييس النفسية تعد أيضًا من هذا النوع؛ نظرًا لأنها تعتمد على مقارنة أداء الفرد بأداء أقرانه فى الجوانب العقلية والوجدانية والمزاجية. أى أن الاهتمام يكون منصبًا على ترتيب الأفراد بعضهم بالنسبة للبعض الآخر من حيث درجاتهم فى الاختبار، ويكون للدرجة تفسيرًا فقط فى ضوء معيار Group Norm يعتمد على متوسط أداء الجماعة التى ينتمى إليها. وسوف نوضح هذا المفهوم بالتفصيل فى الفصل السادس.

خلاصة

تعتمد الدراسة العلمية للسلوك الإنسانى على أساليب القياس والتقويم التربوى والنفسى، كما ينبغى أن تستند القرارات المتعلقة بالأفراد، مثل قرارات الانتقاء والتصنيف، والفحص والتشخيص، والتسكين والترفيع، والتوجيه والإرشاد، وكذلك تقويم المدرسة والمناهج والبرامج والمشروعات التربوية المستحدثة إلى هذه الأساليب. ويتضمن مفهوم القياس عملية جمع المعلومات وتنظيمها، وكذلك نتائج هذه العملية. ونظراً لأن الظواهر السلوكية تتميز بالتعقد والتشابك وتعدد المتغيرات، فإن قياسها يتطلب ابتكار طرق وأساليب وأدوات قياس متنوعة تتميز بخصائص محددة. وقياس السمات الإنسانية يتطلب معرفة طبيعة هذه السمات ومستويات قياسها والقواعد والإجراءات التى تكشف عنها.

لذلك فإن مدخل السمات في القياس التربوى والنفسى يُوضح كيفية تعريف السمات المراد قياسها تعريفًا إجرائيًا، ويحدد الشروط التي يجب أن تتوافر فيها، وخصائصها الأساسية التى تيسر عمليات قياس السلوك الذى يُستدل منه عليها. وبالطبع تتباين أنواع هذه السمات، حيث إن هناك سمات مورفولوجية، وسمات فسيولوجية، وسمات عقلية معرفية، وسمات وجدانية، وسمات مزاجية، وسمات نفسية حركية، وغير ذلك. ويمكن قياس هذه السمات بأدوات متعددة ينبغى أن تفى بشروط معينة، مثل التقنين، والموضوعية، وتمثيلها للسلوك المراد قياسه، وتحديد معيار جماعة مرجعية، أو محك أداء مقبول. ونظراً لتعدد هذه الأدوات، فإنه يمكن تصنيفها من حيث الغرض من القياس، ونوع السمة المراد قياسها، وطريقة التطبيق، والمحتوى، وأسلوب الاداء، وطريقة التصحيح، وكيفية البناء، وتفسير الدرجات المستمدة منها.

الفصل الثانى

مشكلات القياس التربوى والنفسى وشروطه العلمية وأخلاقياته

* مقدمة
* بعض المشكلات الأساسية فى القياس والتقويم فى الوطن العربى
* بعض الأخطاء الناجمة عن إساءة استخدام الاختبارات والمقاييس
* الشروط العلمية لانتقاء الاختبارات والمقاييس
* الشروط العلمية لتفسير درجات الاختبارات وتقييم نتائجها
* بعض ضوابط وأخلاقيات استخدام الاختبارات النفسية والتربوية

مقدمة :

اتضح لنا من الفصل الأول أن القياس التربوى والنفسى أكثر صعوبة من القياس الفيزيائى، وربما أقل دقة بسبب طبيعة الظواهر السلوكية وتعقدها وتشابكها. إذ تتأثر هذه الظواهر بالعديد من المتغيرات الصريحة والضمنية مما يجعل عزل المتغير أو السمة المراد قياسها عن غيره من المتغيرات من المشكلات التى تتطلب جهوداً بحثية متواصلة.

لذلك فإن استخدام الاختبارات والمقاييس التربوية والنفسية ليس ببساطة استخدام مقاييس الظواهر الطبيعية، وإنما يتطلب دراية تامة بأساسيات علم القياس وأساليبه وأدواته وتطبيقاته، والإفادة الواجبة من التطورات المتسارعة التى حدثت فى مختلف مجالاته، ومعرفة مشكلاته وكيفية التغلب عليها بأساليب مناسبة.

كما يتطلب أطراً بشرية مؤهلة تأهيلا علميا ومهنيا فاعلا، ومتدربة تدربًا سليمًا على استخدام أدوات القياس المتنوعة، من مثل الاختبارات،والاستبيانات،والقوائم، وبطاقات الملاحظة،والبروتوكولات وغيرها استخدامًا واعيًا مستبصراً، وتفسير نتائجها تفسيراً مناسبًا يتفق وطبيعة أدوات القياس المستخدمة. كذلك يتطلب معرفة ما يترتب على القرارات التى تتخذ فى ضوء نتائج القياس من مشكلات اجتماعية وأخلاقية. لذلك فإن الالتزام بالمعايير والمبادئ الأخلاقية التى تحكم ممارسات استخدام الاختبارات والمقاييس التربوية والنفسية التى أقرتها المؤسسات المعنية المتخصصة يصبح أمرًا أساسيًا وضروريًا.

لذلك فإننا سنعرض فى هذا الفصل بعض مشكلات القياس التربوى والنفسى، وبعض الأخطاء التى يمكن أن تنجم عن إساءة استخدام الاختبارات والمقاييس، والشروط والضوابط التى تحكم عمليات انتقاء وتطبيق واستخدام أدوات القياس والتقويم فى مختلف المجالات التطبيقية التى أوضحناها فى الفصل الأول، وكذلك أخلاقيات استخدام هذه الأدوات.

بعض المشكلات الأساسية في القياس والتقويم في الوطن العربي

إن الخدمات التي تقدمها مختلف المؤسسات التربوية والنفسية تستهدف مساعدة الأفراد في الإفادة من قدراتهم وإمكاناتهم وطاقاتهم إلى أقصى درجة ممكنة، وتنمية هذه الطاقات، وتمكينهم من وضع أفضل الخطط التي تسهم في تحقيق الأهداف التي يسعون إليها، والتغلب على مشكلاتهم سواء التعليمية أو الحياتية. ويتطلب تقديم هذه الخدمات على أفضل وجه الاستعانة بجميع الوسائل الفنية والاختبارات والمقاييس التربوية والنفسية.

وتُقَدَّم هذه الخدمات من خلال المؤسسات التعليمية والتدريبية ومراكز العلاج المتخصصة وغيرها. ويشرف على الخدمات النفسية عادة إدارات أو مراقبات لهذه الخدمات تتضمن أقسامًا مختلفة من بينها قسم للاختبارات والمقاييس أو القياس والتقويم يختص بإعداد وتقنين أدوات القياس الجماعية والفردية في المجالات النفسية والتربوية والمهنية المتصلة بعمليات الخدمات النفسية والتربوية، وكذلك إعداد وتقنين أساليب جمع البيانات التي تحتاجها عمليات الإرشاد النفسي والتربوي، وتوفير المقاييس النفسية التي تستخدم في دراسة وتشخيص الحالات.

غير أن هناك - في رأينا - عددًا من المشكلات المتعلقة بالقياس والتقويم التربوي والنفسي تعوق فاعلية هذه الخدمات من أهمها :

إسناد مهام القياس النفسي والتربوي إلى غير الاختصاصيين:

فالأفراد الذين يوكل إليهم القيام بمهام القياس النفسي والعمليات الاختبارية تتباين مؤهلاتهم وأساليب إعدادهم ومدى تدريبهم أو درجة خبرتهم، ويزاولون أحيانًا هذه المهام التي تؤثر تأثيرًا بالغًا في حياة الأفراد دون تصريح رسمي. إذ أحيانًا يقوم بعض أمناء أو مساعدي المختبرات النفسية بانتقاء الاختبارات وتطبيقها، وتقديم نتائجها للمفحوصين، بل وأحيانًا يقوم بتطبيق أو استخدام الاختبارات الفردية مثل مقياس بينيه Binet أو ويكسلر Wechsler في الذكاء، أو اختبار الرورشاخ Rorschach، أو اختبار تفهم الموضوع TAT في الشخصية مساعدي باحثين أو أخصائيين نفسيين غير مدربين تدريبًا كافيًا على استخدام هذه الأدوات وتفسير نتائجها في ضوء المحددات المختلفة للاستجابات.

وهذا الأمر يتطلب من المؤسسات المعنية بالخدمات النفسية أن تقوم بإعداد مؤشرات عن مؤهلات العاملين في هذا المجال ونوعية تدريبهم ومدى خبرتهم للاسترشاد بها في تحديد كفايات هؤلاء العاملين. ونحن لا نقترح هنا ضرورة حصول القائمين بأمور التقويم ومن يسند إليهم مهامه على درجة الدكتوراه في علم النفس،

وإنما يمكن العمل على تطوير برامج للدراسات العليا فى الجامعات والمعاهد العليا تمنح الملتحقين بها دبلومات تخصصية فى مختلف المجالات التطبيقية لعلم النفس على أن يكون القياس والتقويم النفسى والتربوى أحد هذه المجالات الرئيسية، وأن يكون التركيز فيه على الجوانب التطبيقية للاختبارات والمقاييس كما هو الحال فى بعض الكليات. وكذلك يمكن أن يسند الأمر فى هذا المجال إلى اختصاصيين نفسيين حاصلين على درجة الماجستير فى علم النفس إضافة إلى حصولهم على دورات تدريبية تختلف باختلاف أنواع الاختبارات، وأن يلتزموا بالمستويات والمعايير والضوابط المهنية فى هذا المجال كما سنعرض فيما بعد.

خبرة الاختصاصيين النفسيين المحدودة بنظريات القياس النفسى والتربوى والأسس التى يقوم عليها:

فالقياس والتقويم النفسى والتربوى علم وفن له أسسه ونظرياته وأصوله ومهاراته، ويجب ألا يتصدى له إلا من ملك ناصية علومه وتدرب على فنونه، فتفسير نتائج الاختبارات دون دراية كافية بأسسها النظرية يجعل الاختصاصى النفسى كالفنى وليس كالاختصاصى المهنى؛ لأنه بذلك يقوم بتنفيذ ما يوصى به بدلا من تسلحه بالمعارف والمهارات التى تمكنه من انتقاء أو بناء أدوات قياس وأساليب تقويم تناسب أغراضًا معينة وأفرادًا معينين.

وترتبط هذه المشكلة ارتباطًا وثيقًا بالمشكلة السابقة وهى إسناد هذه المهام إلى أفراد ينقصهم الإعداد والخبرة. فمثلاً يوكل أمر بناء الاختبارات التحصيلية والتربوية فى المؤسسات التعليمية إلى معلمين لم يدرسوا علم القياس النفسى والتربوى دراسة وافية، ولم ينالوا قدرًا كافيًا من التدرب والخبرة فى هذا المجال بما يمكنهم من إعداد هذه الأدوات إعدادًا علميًا سليمًا، وتفسير نتائجها فى ضوء المحكات والمعايير المحددة. لذلك أوصى اجتماع الخبراء فى الحلقة الدراسية الإقليمية لليونسكو حول تنويع التعليم الثانوى وربطه بالعمل المنتج فى الدول العربية فى الكويت عام ١٩٨٧ وغيره من المؤتمرات العربية المتخصصة بضرورة تطوير مقررات القياس والتقويم النفسى والتربوى فى برامج إعداد المعلمين بحيث تراعى الأساليب والتقنيات الحديثة فى هذا المجال.

قلة الاختبارات المقننة واختبارات الكفايات فى البلاد العربية:

فقيام الاختصاصى النفسى بالمهام الموكلة إليه بأسلوب علمى موضوعى يتطلب توافر أدوات قياس مقننة لمختلف جوانب شخصية الفرد ومقاييس لكفايات الأداء فى مختلف الأعمال والمجالات. غير أن المتاح من هذه الأدوات يُعد قليلاً إذا ما قورن

بغيرها من البلاد، وهي في معظمها مقتبسة من ثقافات أخرى تختلف عن ثقافاتنا مما يكون له تأثير معين على الأداة بعد تعريبها وتقنينها على البيئة المحلية.

فالأصل في المقياس أن يتفادى ما يشوبه، ولكن الأثر الثقافي لا يزال واضحًا في الكثير من المقاييس المعتمدة في بلادنا، ومن هذه الزاوية يكون الضعف أو التحيز الذي سبق أن أشرنا إليه. وهذا التحيز لا يظهر في التقليل من شأن الأفراد فحسب بل يظهر كذلك في منح آخرين درجات أكثر مما يستحقون، وكذلك لا يكون التحيز ضمن المجتمع الواحد فحسب، بل يتجاوز ذلك إلى مجتمع آخر يأخذ بالمقياس دون توفير معايير مناسبة لهذا المجتمع.

فالمقاييس ينبغي أن يجرى إعدادها اعتمادًا على عينة كافية وضمن شروط علمية محددة. وأن يكون قادرًا في البيئة التي وضع لأفرادها أن يقيس ما وضع لقياسه. ومن هذه الزاوية يكون من الخطر اعتماده في بيئة أخرى تختلف كثيرًا عن السابقة دون إجراء التعديلات التي تقتضيها عمليات التعيير أو التقنين ليتناسب مع البيئة الجديدة، وهذا أمر يحتاج إلى عناية فائقة.

كما أن الأصل في المقياس أن تكون لغته مناسبة لمن يفحص به. ولكن وجود لغة يومية، أي لغة دارجة تسود في مجتمع ما وتختلف من جهات عن اللغة التي تظهر في الكتب والصحف اليومية والدوريات تطرح مسألة حادة وخاصة في البلاد العربية.

وقد حاولت بعض هذه البلاد كتابة عبارات أو أسئلة أو بنود الاختبارات والمقاييس النفسية باللغة اليومية «الدارجة» التي يفهمها عامة الناس. غير أن هذا يتطلب وقفة تأمل حول الآثار والمضامين الثقافية الناجمة عن التوسع في هذا الاتجاه في مجال القياس النفسي وبناء الاختبارات.

ومن المشكلات الهامة الأخرى في هذا المجال استخدام الاختبارات التحصيلية التي يكتبها المعلمون في أغراض متعددة مثل الانتقاء، والفحص، والترفيع والتوجيه، وعدم إيلاء الاختبارات التشخيصية والبنائية أهمية تذكر. فنحن نعاني في بلادنا من نقص شديد في هذه الأنواع المهمة من الاختبارات التي تحدد أوجه القصور ومواطن القوة في تحصيل الطلاب بالنسبة لمجموعة محددة من المعارف والمهارات. وترشد العملية التعليمية وتصوب مسارها نحو تحقيق أهدافها المرجوة. كما نعاني من قلة الاختبارات التحصيلية المقننة في مختلف المواد الدراسية والتي تفيد في تحديد المركز النسبي للطالب بين زملائه في محتوى أو محتويات دراسية معينة.

وعلى الرغم من الجهود التى أفادت من التطورات التى حدثت فى هذا المجال منذ العقدين الماضيين إلا أنه ينبغى تجميع وتنظيم هذه الجهود وتكثيفها فى مؤسسات متخصصة.

ويمكن أن يتم ذلك بتحديد الجهات أو المؤسسات التى يُسند إليها بناء الاختبارات والمقاييس وتزويدها بالخبراء والاختصاصيين والتقنيات المناسبة. كما ينبغى حصر وتصنيف ما هو متوافر فى البيئة المحلية من اختبارات نفسية وتربوية ومراجعتها بواسطة خبراء القياس والتقويم لتحديد الصالح منها، وإعادة النظر فى الاختبارات التى تتطلب ذلك.

بعض الأخطاء الناجمة عن إساءة استخدام الاختبارات والمقاييس النفسية والتربوية:

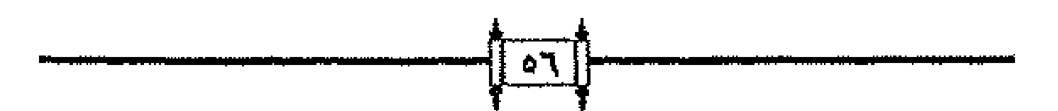
على الرغم من أن الاختبارات والمقاييس النفسية والتربوية تعد أحد مصادر جمع المعلومات؛ إلا أنها بما تتميز به من خصائص مهمة، من مثل الموضوعية والتقنين والمعايير والثبات والصدق، تعد من أهم هذه المصادر (Cronbach , 1970) .غير أن سوء استخدام هذه الأدوات من قبل أفراد غير متخصصين أو غير متدربين تدربًا عاليًا، وغير مطلعين على التطورات المتنامية فى مجال القياس والتقويم النفسى والتربوى، يؤدى إلى أضرار بالغة بمن تطبق عليهم هذه الأدوات، ويفقدها فعاليتها، ويحيد بها عن تحقيق الأغراض الأساسية منها، بل وتصبح مصدراً من مصادر الشك والنقد من مختلف فئات المجتمع (Aiken, 1985) .

وافتقارنا فى الوطن العربى إلى أسس ومعايير وضوابط واضحة ومحددة نابعة من ممارساتنا المهنية الصحيحة فى مجال استخدام أدوات القياس النفسى والتربوى، ومسترشدة بواقعنا ومبادئنا الخلقية والاجتماعية والثقافية، يؤدى إلى وقوع الاختصاصيين النفسيين فى أخطاء ناجمة عن سوء استخدام هذه الأدوات وعن تقديم نتائجها إلى غير المتخصصين أو إلى المفحوصين.

فالممارسات الخاطئة لاستخدام الاختبارات والمقاييس النفسية لا تؤدى فقط إلى قرارات غير صائبة، وإنما لها آثارًا ضارة بالمفحوص وانعكاسات اجتماعية سيئة.

ولا نتعدى الحقيقة إذا قلنا أن هذه الممارسات الخاطئة كانت سببًا أساسيًا فى الهجوم الشديد والنقد المستمر للاختبارات والمقاييس النفسية ذاتها.

ومن أهم هذه الممارسات الخاطئة وما يترتب عليها من مشكلات ما يلى:



تحيز المسئولين عن البرامج الاختبارية : فبعض المسئولين عن البرامج الاختبارية أو التقويمية لا يراعون تباين الأفراد المختبرين في خلفياتهم الثقافية وخصائصهم الشخصية المميزة عند انتقاء الاختبارات والمقاييس النفسية واستخدامها، وهذا يؤثر بلا شك في نتائج الاختبارات، وبالتالي يقلل من شأن هؤلاء الأفراد.

الثقة التامة في التنبؤ باستخدام الاختبارات : إن الثقة التامة في التنبؤ باستخدام الاختبارات والمقاييس المقننة يعد من الممارسات الخاطئة. فالتنبؤ بالسلوك المستقبلي باستخدام مختلف أنواع الاختبارات والمقاييس ينبغي أن يكون في حدود معينة من الثقة. إذ إن هذا التنبؤ يرتبط بعوامل متعددة تتعلق بالفرد المختبر، والسلوك المراد التنبؤ به، والمدة الزمنية للتنبؤ، وخصائص المقياس المحك الذي يقدر في ضوئه فاعلية التنبؤ، وإذا قلت الثقة في التنبؤ بدرجة ملحوظة يصبح استخدام هذه الأدوات مضللاً، وبخاصة إذا كان القائم باستخدامها شخص غير مدرب أو قليل الخبرة.

كما أن استخدام الاختبارات في أغراض الانتقاء والتصنيف التي لم يتم التحقق من صدقها أو اتُّبع في سبيل التحقق من ذلك إجراءات واهنة يعد أيضًا من الممارسات الخاطئة. فمشكلة صدق الاختبارات والمقاييس النفسية تتضمن في واقع الأمر قضايا متعددة ولا تقتصر فقط على مجرد تقدير خاصية من خصائص الاختبار أو المقياس.

فالاستخدام المناسب للاختبارات يتضمن العديد من المسئوليات التي يعتمد بعضها على درجة ثبات وصدق الاختبار، ويعتمد البعض الآخر على الظروف المختلفة المتعلقة بتطبيقات الاختبار.

الاستخدام الروتيني لدرجات الاختبارات :

فالدرجات المستمدة من الاختبارات تزودنا بقاعدة من البيانات تفيد في تصنيف الأفراد تصنيفًا اعتباريًا. وهذا يعني أن هذه الدرجات ليست قياسات ثابتة ولكنها قيم تقديرية لسمات إنسانية، وبالتالي فهي تحتوي على أخطاء معاينات مختلفة (Anastasi, 1982). غير أن هذه الاختبارات تستخدم عادة بطريقة روتينية لا تتسم بالمرونة الكافية التي تأخذ مختلف المتغيرات المؤثرة بعين الاعتبار. وهذه الممارسة الخاطئة تؤدي في بعض الأحيان إلى تصنيف التلاميذ مثلاً في مجموعات بحسب قدراتهم اعتمادًا على اختبارات أو مقاييس ربما يصعب تبرير استخدامها، أو استخدام درجات فاصلة Cut - off Scores في الاختبارات واعتبارها محكات شبه ثابتة لانتقاء الأفراد في مجالات عمل معينة في المؤسسات الصناعية.

افتراض أن الاختبارات العقلية تقيس خصائص موروثة :

فأحيانًا يفسر بعض مستخدمي الاختبارات درجاتها على أنها قياسات لخصائص موروثة، ومثال ذلك سمة الذكاء. وهذا التفسير الخاطئ له آثار غير مرغوبة، إذ إنه

يؤدى إلى اعتبار أن القدرات الإنسانية ثابتة أو غير قابلة للنمو، بل وربما يؤدى إلى النظر إلى الإنسان نظرة تصنيفية، كأن نقول أن هذا متخلف عقليًا وهذا عبقرى وما شابه ذلك، ويرى بعض النقاد أن هذه نظرة غير إنسانية.

أخطاء ناجمة عن تقديم الاختبارات إلى غير المتخصصين أو إلى المفحوصين:

هناك ممارسات خاطئة أيضًا تنجم عن تقديم نتائج الاختبارات والمقاييس النفسية إلى أفراد غير متخصصين أو إلى المفحوصين. ومن بين هذه الأخطاء ما يلى:

مسايرة التوقعات لنتائج الاختبارات: فقد بينت دراسة روزنثال وجاكوبسون (Rosenthal& Jacobson, 1970) أنه إذا أعطى للمعلمين معلومات افتراضية - أى ليست حقيقة - عن قدرات تلاميذهم فإن التحصيل الفعلى لهؤلاء التلاميذ يتفق وتوقعات المعلمين عنهم. فالتلاميذ الذين توقع المعلم أن يكون تحصيلهم منخفضًا كان تحصيلهم كذلك بالفعل. والعكس كان صحيحًا. لذلك فإن الذين ينتقدون اختبارات الذكاء يرون أن هذه الاختبارات تؤثر فى أداء الأفراد تأثيرًا غير مرغوب فيه. إذ يرون أن المعلمين مثلاً ربما ينظرون إلى التلاميذ مرتفعى الذكاء نظرة تختلف عن نظرتهم للتلاميذ الأقل ذكاء فى ضوء ما يتوقعونه من كل منهم.

تأثر مفهوم الذات لدى الفرد بنتائج الاختبارات: فالأخصائى النفسى غير المتمرس ربما يقدم نتائج الاختبارات والمقاييس النفسية للمفحوص بطريقة تؤثر سلبيًا فى مفهومه عن ذاته. إذ ربما يؤدى تعميم المفحوص لنتائج الاختبار إلى استنتاج أشياء عن نفسه ليس لها ما يبررها، فالشاب الصغير الذى يحاول جاهدًا أن يكوّن مفهومًا إيجابيًا عن ذاته ربما يتأثر هذا المفهوم لديه ويقل مستوى طموحه إذا علم أن درجاته منخفضة فى بعض الاختبارات. فالتقارير المقتضبة لنتائج الاختبارات وما يبنى عليها من توصيات تقدم للأفراد فى بعض مواقف الإرشاد النفسى تؤثر بلا شك تأثيرًا سلبيًا فى نظرتهم إلى أنفسهم، وفى تشكيل اتجاهاتهم نحو ذواتهم.

والآن ماذا بعد...؟ هل إساءة استخدام الاختبارات والمقاييس النفسية والتربوية تعنى المناداة بعدم استخدامها فى مجالات الخدمات النفسية والتربوية؟

إن كرونباك (Cronbach, 1970) يغنينا عن إجابة هذا السؤال بقوله: «إن السياسات الحكيمة لا تدافع عن الاختبارات ولا تهاجمها، وإنما ترشد كيفية استخدامها بما يتفق والأهداف المرجوة منها». وهو بذلك يؤكد أن فاعلية أدوات القياس النفسى والتربوى التى يعتنى بتصميمها وإعدادها يتوقف تحقيقها لأهدافها على نوعية استخدامها فى مختلف مجالات الخدمات النفسية والتربوية وكفاءة وخبرة القائمين بها.

الشروط العلمية لانتقاء الاختبارات والمقاييس النفسية والتربوية :

إن إجراء الاختبارات يعد نوعًا من التواصل والتفاعل الاجتماعي بين طرفين أو أكثر، وهذا يتطلب اتجاهًا أخلاقيًا مسئولا من جانب الفاحص، والرغبة في التعاون من جانب المفحوص، فجميع أشكال التفاعل الاجتماعي المثمر تتطلب الثقة والاحترام المتبادل بين الأطراف المتفاعلة.

فإذا كان المفحوص واثقًا من النفع الذي سيعود عليه نتيجة للوقت والجهد الذي يبذله في أداء الاختبارات التي يجريها عليه فاحص متخصص، فإنه سوف يحاول جاهدًا أن يتفاعل معه تفاعلاً إيجابيًا، ويتبع تعليماته؛ لذلك تطبق على المفحوص بهدف مساعدته على فهم نفسه وصنع قراراته.

فكثير من الممارسين النفسيين يعتمدون في عملهم على عدد محدود من الاختبارات دون سواها، إما لأنها مألوفة لديهم واعتادوا على استخدامها، أو لأنها من إعداد أشخاص من ذوي مكانة علمية متميزة. ولكن هذه الأسباب لا يجب أن تكون مانعًا لهم من البحث عن اختبارات ومقاييس جديدة متنوعة. ويمكنهم تتبع هذه الأدوات الجديدة في المطبوعات التي تصدرها دور النشر ومؤسساتها المختلفة، وفي الدوريات العلمية المتخصصة.

غير أن هذا يتطلب توافر مهارات معينة لدى الأخصائيين النفسيين تساعدهم في الحكم على هذه الأدوات، ومدى ملاءمتها للغرض الذي سوف تستخدم من أجله. كما يتطلب توفير سبل توثيق الروابط فيما بينهم لتبادل المعلومات والخبرات في هذا المجال وغيره من المجالات المتعلقة بشئون مهنتهم.

وفيما يلي بعض الأسس العلمية والاعتبارات الفنية التي يجب مراعاتها في انتقاء الاختبارات والمقاييس التي تستخدم في المجالات المختلفة للخدمات النفسية:

توافر خصائص الاختبار أو المقياس الجيد : فعلى الرغم من تنوع أدوات القياس إلا أنه يجب أن يتوافر فيها خصائص معينة يصعب بدونها الاعتماد عليها في اتخاذ قرارات صائبة تتعلق بالأفراد. ومن أهم هذه الخصائص الموضوعية والثبات والصدق ووجود معايير مستمدة من البيئة التي يستخدم فيها الاختبار. ويمكن أن يحصل مستخدم الاختبار أو المقياس على بيانات تتعلق بهذه الخصائص من دليل الاختبار. وهذا يجعلنا نؤكد أهمية تضمين بيانات وافية ومعلومات كافية تتعلق بهذه الخصائص في أدلة الاختبارات، وأن يكون مستخدم الاختبار على درجة عالية من الكفاية في تقييم هذه الخصائص ومعرفة دلالتها لكي يتمكن من انتقاء أداة القياس الصالحة للغرض

المعين. ولا يجوز انتقاء اختبار أو مقياس من مجرد النظر إلى عنوانه أو اسمه بل لابد من فحص محتوى الاختبار فحصاً مستنيراً والاطلاع على خصائصه السيكومترية وتحديد دلالتها، والتأكد من ملاءمة معاييره للفرد أو العينة المستهدفة، والتحقق من مدى صلاحية هذه المعايير، وسوف نناقش هذه الخصائص مناقشة مستفيضة فى الفصول الخمسة التالية.

مناسبة الاختبار للمفحوص : فانتقاء الاختبار المناسب للمفحوص يعد أمراً على جانب كبير من الأهمية، فلا يجوز أن تطبق الاختبارات التى تصلح للراشدين على تلاميذ المدارس الابتدائية أو المتوسطة، حيث تكون غير مناسبة للمستوى النمائى لهم، كما أن معاييرها مستمدة من عينات تختلف عن عينات هؤلاء التلاميذ. لذلك يجب أن تشير أدلة الاختبارات إشارة واضحة إلى طبيعة عينة تقنين الاختبار وخصائصها، من حيث العمر والنوع والمستوى التعليمى والاجتماعى وغير ذلك لكى يسترشد بها الفاحص فى انتقاء الاختبار المناسب لفرد أو مجموعة معينة. وينبغى مراعاة أن وجود مشكلات سلوكية، أو عجز بصر أو ضعف سمع، أو حرمان ثقافى، أو صعوبات فى اللغة تجعل الاختبار غير مناسب للمفحوص، ويجب عدم استخدام الاختبار فى هذه الحالات.

مناسبة الاختبار للغرض الذى يقيسه : فعلى الرغم من تنوع وتباين استخدامات الاختبارات والمقاييس النفسية والتربوية، إلا أن الهدف الرئيسى منها جمع بيانات ومعلومات تفيد فى اتخاذ قرارات معينة تتعلق بالأفراد أو الجماعات بأقل قدر ممكن من الخطأ أو الصدفة أو المخاطرة. غير أن الاختبار أو المقياس الذى يفيد فى موقف معين قد لا يفيد أو لا يناسب موقفاً آخر، فاستخدام الاختصاص النفسى أداة قياس مناسبة لموقف معين أو صالحة لغرض معين يساعده بالطبع فى اتخاذ قرارات أفضل من تلك التى يستخدمها دون استخدام تلك الأداة.

لذلك يجب تحديد الغرض من عملية القياس أو العملية الاختبارية تحديداً واضحاً وانتقاء أفضل الأدوات التى تحقق هذا الغرض. وهذا يتطلب بالطبع توافر عدد كبير من الاختبارات والمقاييس الجيدة التى تقيس مختلف السمات الإنسانية حتى يتمكن الفاحص من انتقاء أكثرها ملاءمة للغرض من القياس. وهنا تبرز أهمية تشجيع حركة القياس النفسى والتربوى وبناء الاختبارات بأنواعها المختلفة فى المنطقة العربية من خلال مؤسسات متخصصة.

الشروط العلمية لتطبيق الإختبارات وتصحيحها :

بعد انتقاء أدوات القياس التى تحقق بدرجة أفضل أهداف العملية الاختبارية وتناسب المفحوص أو المفحوصين، تجرى عملية تطبيق الأدوات.

وقد يظن البعض أن تطبيق الاختبارات عملية آلية بسيطة لا تحتاج إلى توافر مهارات معينة فيمن يقوم بها، غير أننا إذا تذكرنا أن تطبيق الاختبارات يُعد عملية تفاعل اجتماعى كما سبق أن أشرنا، فإننا سرعان ما ندرك تعدد المتغيرات والظروف النفسية والفيزيائية المتعلقة بالموقف الاختبارى التى تؤثر فى أداء المفحوص فى الاختبارات وفى نتائج العملية الاختبارية. وإذا لم يكن القائم بتطبيق الاختبارات شخصًا مؤهلاً أو مدربًا للقيام بهذا العمل فإنه ربما لا يلتفت إلى هذه المتغيرات.

فالقلق والتوتر،وانخفاض مستوى الدافعية، وتشتت الانتباه، والضعف الحسى، وعدم الألفة بالاختبارات النفسية، والكذب وتزييف الاستجابات، والتخمين العشوائى، والاتجاه نحو الاختبار ونحو الفاحص، واحتمال عدم فهم المفحوص للتعليمات بسبب ضعف البصر أو انخفاض المستوى اللغوى ، والظروف الفيزيائية غير الملائمة، تؤثر تأثيرًا ملحوظًا فى أداء المختبرين. ونحن عادة نطمح فى أن تطبق الاختبارات فى ظروف مقننة، ولكن هذا لا يتحقق بدرجة تامة فى كثير من المواقف الاختبارية.

لذلك يجب أن تشتمل كراسة تعليمات الاختبارات أو ملحقاتها على توجيهات وتعليمات واضحة ومحددة عن كيفية ضبط هذه المتغيرات، وأن يقوم الفاحص بدراستها دراسة متأنية قبل تطبيق الاختبار، وكذلك الاهتمام بتهيئة المفحوصين لأداء الاختبار من خلال مقابلات فردية أو فى مجموعات صغيرة لإزالة أو تخفيف مصادر القلق والتخوف وغيرها من الانفعالات السالبة، ومحاولة رفع مستوى الدافعية بدرجة تسمح للمفحوص ببذل الجهد المطلوب لأداء الاختبار، ومن ثم ينظم الموقف الاختبارى فى ضوء هذا الإعداد قبل اليوم المحدد لتطبيق الاختبار.

كما يجب أن تشير كراسة تعليمات الاختبار إلى الشروط اللازم توافرها فيمن سيقوم بتطبيق الاختبار.

وسوف نعرض فيما يلى بعض الشروط العلمية والتوجهات الفنية التى يجب مراعاتها عند تطبيق الاختبارات وتصحيحها:

(١) ينبغى على من يقوم بتطبيق اختبار أو مقياس معين أن يتبع الإجراءات المقننة المذكورة فى كراسة تعليمات الاختبار بعناية تامة، فهذه الإجراءات المقننة ضرورية إذا كان الأخصائى النفسى يود التوصل إلى قرارات تتعلق بالمختبر اعتمادًا على

درجات الاختبار. ولا يجوز الحيد عن هذه التعليمات أو اختصارها أو تعديلها. ومن الضرورى تدريب القائمين بتطبيق الاختبارات على أداء متطلبات هذا العمل بكفاءة وفعالية. وبالطبع تختلف هذه المتطلبات باختلاف نوع الاختبار، إذ إن بعض الاختبارات وبخاصة الفردية التطبيق تحتاج إلى إشراف مباشر من الأخصائى (Anastasi,1982).

(٢) ينبغى تهيئة ظروف متسقة لتطبيق الاختبار تساعد على ضبط متغيرات الموقف الاختبارى بقدر الإمكان، فمثلا لا يجوز أن يكون هناك اختلاف كبير فى درجة الحرارة أو الرطوبة أو الضوضاء، وما يمكن أن يؤدى إلى تشتت انتباه المفحوصين فى الغرفة أو القاعة المخصصة لإجراء الاختبار، وأن يجرى فحص دورى على الأجهزة والمواد أو الأدوات الاختبارية، من مثل ساعات التوقيت والأجهزة الإلكترونية واليدوية المختلفة للتأكد من سلامتها وتقنينها.

كما يجب التأكد من أن المفحوص قد فهم تعليمات الاختبار ونوع الاستجابة المطلوبة منه، وأن يهيئ الفاحص بيئة إنسانية يسودها المودة والحرص والمسئولية، وبخاصة إذا كان هناك اختلاف بين الفاحص والمفحوص فى النوع أو المركز الاجتماعى حتى يشعر المفحوص بالاحترام والثقة بالنفس.

(٣) إن الأخصائى النفسى مسئول عن دقة تصحيح الاختبارات ومراجعة عمليات التصحيح وتسجيل النتائج؛ لذلك يجب أن يطمئن على أن القائم بهذه العمليات يتحرى الدقة التامة، ويبذل جهده فى التحقق منها.

وإذا كان التصحيح يتطلب تقدير درجات اعتماداً على أحكام خبراء، فإنه يجب على الأخصائى تحديد ثبات هذه التقديرات للتأكد من اتساقها، إذ إنه ليس من العدل اتخاذ قرارات تمس حياة الأفراد اعتماداً على درجات غير صحيحة أو غير متسقة. وإذا أجرى التصحيح باستخدام أجهزة إلكترونية يجب على الأخصائى النفسى التحقق من سلامة هذه الأجهزة ودقتها، وأن يتحقق من النتائج بإعادة التصحيح يدويًا لبعض أجزاء الاختبار، أو من خلال نظام معين يضمن الضبط الكيفى لعملية التصحيح.

وعلى الرغم من تطور الأساليب الآلية فى تصحيح الاختبارات، وإعداد التقارير، إلا أن العنصر البشرى سوف يظل هو العامل الأساسى فى جميع مراحل العملية الاختبارية؛ لذلك يجب الاهتمام بانتقاء الأفراد الذين يوكل إليهم تصحيح الاختبارات والمقاييس النفسية بحيث تتوافر فيهم صفات المثابرة والدقة التامة واللماحية، ويستطيعون إجراء العمليات الحسابية بسرعة ودقة. وأن تُنظم لهم دورات تدريبية تحت إشراف أخصائيين نفسيين من ذوى الكفاءة العالية.

الشروط العلمية لتفسير درجات الاختبار وتقديم نتائجها :

إن المرحلة التالية لتطبيق الاختبارات وتصحيحها تتعلق بتفسير درجاتها وكتابة تقارير النتائج، وتقديمها للمفحوص أو الجهة المعنية، وتُعد هذه المرحلة من المراحل المهمة في العملية الاختبارية، إذ يترتب عليها اتخاذ قرارات بشأن الأفراد أو المؤسسات.

وعلى الرغم من أن بعض مؤسسات العمل تحيط نتائج الاختبارات النفسية التي تجريها على المتقدمين للوظائف الشاغرة بها سرية تامة بحيث لا يعلم المتقدم عن هذه النتائج سوى قبوله أو رفضه بحجة أن هذه المؤسسات ليست بصدد التوجيه المهني، إلا أن هذا يؤدي بالمفحوص إلى الشك في نتائج الاختبارات ووصفها بالتحيز وعدم النفع. فمن حق المفحوص - أو وكيله أو ولي أمره إذا كان المفحوص قاصراً - معرفة نتائج الاختبارات. وتقديم نتائجها يتطلب قدراً كبيراً من الخبرة والمهارة من جانب الاختصاصي النفسي في تطبيق أسس ومفاهيم القياس النفسي والإحصاء، وأساليب الإرشاد النفسي في استخلاص المعلومات التي تقدم للمفحوص.

وفيما يلي بعض الشروط التي يجب مراعاتها في هذا الشأن:

(١) ينبغي العناية بتفسير درجات الاختبارات والمقاييس المقننة في ضوء المعايير الخاصة بها والمدونة في دليل الاختبار أو المقياس، والالتزام بهذه المعايير وعدم الحيد عنها.

(٢) ينبغي تقديم التقارير التي تتضمن درجات الاختبارات لأفراد مؤهلين لتفسيرها واستخدامها استخداما مناسبًا. أما التقارير التي تقدم للآباء أو المعلمين أو المشرفين على العمل، فمن الأفضل أن تتضمن تفسيراً للنتائج بدلا من تقديم الدرجات ذاتها. فكثير من هؤلاء الأفراد ربما لا يستطيعون تفسير معنى الدرجات المعيارية أو المئينيات وغيرها من الدرجات المحوّلة، وأن يشرف الاختصاصي النفسي على هذه التقارير ويكون مستعداً لتقديم العون والمشورة لهؤلاء الأفراد.

(٣) ينبغي أن تتباين أشكال تقارير نتائج الاختبارات بتباين الأفراد أو الجهات التي تقدم إليها هذه النتائج، إذ يجب أن تمكنهم هذه التقارير من فهم تفسير هذه النتائج بيسر وسهولة.

(٤) ينبغي تجنب استخدام كلمات وصفية مثل متخلف عقليًا أو عدواني عند تفسير درجات الاختبارات والمقاييس النفسية، أو كلمات مثل راسب أو غير كفء في الاختبارات التحصيلية واختبارات الكفايات، فهذه الكلمات أو الأوصاف التي تُعبِّر عن

أحكام قيمية تكون عرضة لأخطاء التفسير من جانب الآباء والمعلمين والمفحوص ذاته. لذلك يفضل وصف أو تحديد السلوك الذى يميز الفرد أو الذى يستطيع أو لا يستطيع أداؤه.

(٥) ينبغى الحيطة عند تفسير نسب الذكاء والعمر العقلى ومعايير الفرق الدراسية وما شابه ذلك، إذ إن هذه المعايير يشوبها كثيراً من العيوب التى تؤدى إلى عدم دقة التفسير. وإذا اضطر الاختصاصى النفسى إلى تقديم مثل هذه النتائج إلى مؤسسة معينة، فإنه يُفضل أن يقدِّم التفسير مستخدمًا الدرجات المعيارية أو الميئنيات المناظرة فى ضوء عينة التقنين التى استخدمت فى اشتقاق هذه الدرجات، ويجب أن يشار فى التقرير أيضًا إلى صورة الاختبار المستخدمة وتاريخ إجراء الاختبار، وطبيعة الموقف الاختبارى.

بعض ضوابط وأخلاقيات استخدام الاختبارات والمقاييس النفسية والتربوية:

يتضح مما سبق عرضه من شروط علمية واعتبارات فنية لمختلف مراحل العملية الاختبارية أن استخدام الاختبارات النفسية علم وفن، فالاختبارات هى أدوات قياس تعين الأخصائيين فى مختلف مجالات الخدمات النفسية فى أداء واجباتهم الإنسانية بأسلوب علمى موضوعى، واستخدامها فى تحقيق الأهداف التى أعدت من أجلها يتوقف على نوعية هذا الاستخدام وكفاءة وخبرة العاملين فى هذا المجال.

ولضمان التزام جميع الأطراف المعنية باستخدام الاختبارات والمقاييس النفسية ينبغى أن تكون هناك مجموعة من الضوابط والأخلاقيات التى توجه المسار، وتحدد الممارسات فى هذا المجال، وتضفى على المهنة الصفة الكاملة التى تستوجبها من المجتمع، وترعى حقوق وواجبات القائمين بهذه الخدمات والمستفيدين منها.

ولذلك حرصت الجمعية الأمريكية لعلم النفس (APA) على إصدار دليل متجدد يتضمن الأسس والمبادئ الأخلاقية التى يجب أن يلتزم بها الأخصائيون النفسيون فى أدائهم لمسئولياتهم، ومن بينها المبادئ الأخلاقية المتعلقة باستخدام الاختبارات النفسية وأدوات القياس التشخيصية.

كما أصدرت دليلا للمعايير التى ينبغى توافرها فى هذه الاختبارات، وإرشادات عن إجراءات استخدامها. وكذلك أصدرت الجمعية الأمريكية لشئون الأفراد والتوجيه دليلاً يتضمن المبادئ الأخلاقية فى مجالات الخدمات النفسية متضمنة مجال تطبيق الاختبارات والمقاييس.

وعلى الرغم من أن هذه الضوابط والأخلاقيات يجب أن تنبع من واقع أعضاء المهنة أنفسهم، إلا أننا سوف نقدم بعض التوجيهات التى يمكن الاسترشاد بها فى هذا الشأن.

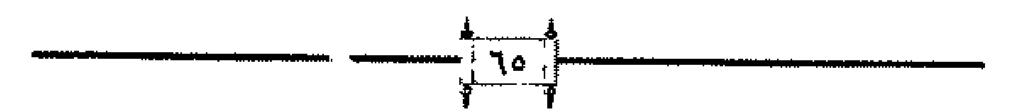
الكفايات اللازم توافرها فيمن يستخدم الاختبارات : إن وضع ضوابط للكفايات اللازم توافرها فيمن يستخدم الاختبارات والمقاييس النفسية تعد خطوة أساسية لتجنب إساءة هذه الاستخدامات، وتختلف بالطبع هذه الكفايات باختلاف نوع الاختبار والهدف منه، ويمكن تصنيف الاختبارات والأدوات التشخيصية والكفايات اللازم توافرها فيمن يستخدمها إلى ثلاثة مستويات (APA,1973) :

المستوى الأول : اختبارات أو أدوات تشخيصية يمكن تطبيقها وتصحيحها وتفسير نتائجها بأسلوب مناسب بالرجوع إلى كراسة تعليمات الاختبار ومراعاة مطالب المؤسسة التى يعمل فيها مستخدم الاختبار، ومن أمثلة ذلك الاختبارات التحصيلية، واختبارات الكفايات المهنية. ويمكن أن يقوم بتطبيق هذه الاختبارات وتفسير نتائجها معلمين ومسئولين عن المؤسسات المختلفة بحيث يكونون ملمين بدرجة معقولة بأساسيات القياس التربوى والنفسى.

المستوى الثانى : اختبارات وأدوات تشخيصية يتطلب استخدامها معرفة أساليب بناء الاختبارات وطرق استخدامها ، ودراسة بعض المقررات فى الإحصاء النفسى، والفروق الفردية، وسيكولوجية التوافق، وسيكولوجية انتقاء الأفراد، والتوجيه المهنى والتربوى. ومن أمثلة ذلك اختبارات الذكاء العام، والاستعدادات الخاصة والاستعدادات الفارقة، ومقاييس الميول، واستبيانات الشخصية. وينبغى فيمن يستخدم هذه الاختبارات والأدوات أن يحقق هذه المتطلبات المذكورة، أو أن يكون مصرحًا له باستخدامها فى إطار مؤسسة تربوية أو صناعية أو إدارة حكومية يعمل فيها، أو أن يكون قد أدى بنجاح دورة تدريبية تتعلق بهذه الاختبارات.

المستوى الثالث : اختبارات وأدوات تشخيصية تتطلب معرفة متعمقة فى القياس النفسى وبعض المجالات النفسية المصاحبة، وكذلك تدريبًا مكثفًا وخبرة عملية تحت إشراف أخصائى إكلينيكى لمدة لا تقل عن عام. ويفضل أن يكون الأفراد الذين يستخدمون هذه الأدوات من الحاصلين على درجة الماجستير فى علم النفس. ومن أمثلة هذه الأدوات: اختبارات الذكاء الفردية، والأساليب الإسقاطية، وبعض مقاييس الشخصية الأخرى.

ويجب على من يستخدم الاختبارات والمقاييس النفسية بعامة أن يكون مزودًا بالمهارات التى تمكنه من انتقاء الاختبارات التى تتفق والهدف من عملية القياس،



وتناسب الفرد الذى تجرى عليه هذه الاختبارات، وأن يكون مطلعًا على الدراسات والبحوث المتعلقة بالاختبار أو الأداة المستخدمة، وكذلك يمكنه تقييم الخصائص السيكومترية للاختبارات، من مثل المعايير، والثبات، والصدق، وأن يكون على دراية بالعوامل النفسية والظروف المختلفة التى تؤثر فى العملية الاختبارية وفى الأداء، ويمكنه التوصل إلى استنتاجات وتوصيات اعتمادًا على درجة أو درجات الاختبار، آخذًا بعين الاعتبار المعلومات الأخرى الهامة المتعلقة بالمفحوص.

حقوق المختبرين : إن من حق الفرد الذى يجرى عليه اختبار أو مقياس معين أن يكون على علم بالهدف من الاختبار وفيم تستخدم نتائجه. فلا يجوز للأخصائى النفسى أن يطلب من المفحوص الإفصاح عن معلومات تخصه أثناء عملية المقابلة الشخصية أو أثناء اختباره أو تقويمه، أو أن يسعى إلى الحصول على مثل هذه المعلومات إلا بعد تعريفه بالهدف من ذلك وأخذ موافقته كتابيًا، أو أخذ موافقة ولى أمره إذا كان قاصرًا.

عدم تداول الاختبارات بين غير المتخصصين فى علم النفس: إن الحفاظ على سرية الاختبارات تعد من مسئولية مؤلفيها وناشريها ويشاركهم فى ذلك الأخصائى النفسى (Sax, 1974). فسرية الاختبارات تعد مشكلة إذا أدت إلى تغيير فى الدرجة التى يحصل عليها الفرد فى الاختبارات دون أن يكون هناك تغيير مقابل فى درجته الحقيقية، غير أن سرية الاختبارات لا تعد مشكلة فى بعض الاختبارات. فمثلاً إذا أجرى اختبار على فرد لقياس درجة تمكنه من مهارة معينة، فإن معرفته السابقة بعينة من أمثلة مشابهة لأسئلة أو تدريبه عليها ربما يوصى به.

ولكن فى كثير من الحالات تؤثر معرفة عينة من الأسئلة أو طريقة تصحيحها تأثيرًا بالغًا فى صدق الاختبار؛ لذلك يجب فى مثل هذه الحالات الحفاظ على الاختبارات فى كبائن مغلقة، وتحديد الأشخاص الذين توكل إليهم مهمة استلام هذه الاختبارات.

وإذا أرسلت الاختبارات إلى مكان آخر عن طريق البريد، فإن الشخص المرسل إليه يجب أن يكون معروفًا وموثوقًا به. والواقع أن آلات التصوير الحديثة والتقنيات المختلفة قد زادت من حدة مشكلة سرية الاختبارات وما يرتبط بها من مخالفات لحقوق الطبع والنشر مما يتطلب اتخاذ الإجراءات التى تضمن منع هذه المخالفات.

نشر وتوزيع الاختبارات : يمكن تلافى كثير من المشكلات الناجمة عن سوء استخدام الاختبارات بتنظيم عملية نشر وتوزيع الاختبارات، فبعض الناشرين يسمحون ببيع الاختبارات والمقاييس النفسية لكل من يطلبها مادام أنه يدفع ثمن شرائها، فى حين

أن عدداً محدوداً منهم يقصر البيع على فئة من المتخصصين. وربما يرجع السبب فى ذلك إلى أن بعض هؤلاء الناشرين لا يدركون الأضرار التى تنجم عن سوء توزيع الاختبارات سواء للأفراد أو المؤسسات.

لذلك فإن هناك حاجة إلى تقييد توزيع أو بيع الاختبارات والمقاييس النفسية، وقصر ذلك على من تتوافر لديهم المهارات أو الكفايات اللازمة لاستخدامها الاستخدام المناسب؛ وذلك لضمان سرية الاختبارات وعدم إساءة استخدامها.

ومن مسئولية القائمين بإعداد الاختبارات أو تأليفها ، ومؤسسات نشرها عدم السماح بتوزيع الاختبارات للاستخدام قبل أن يتم إعدادها فى صورتها النهائية. وإذا سمح بنشر اختبار فى صورته الأولية لأغراض البحث أو الدراسة يجب أن يوضح ذلك صراحة على غلاف الاختبار وفى كراسة تعليماته، وأن يقتصر توزيعه على تلك الأغراض فقط. ويجب أن تشتمل كراسة تعليمات الاختبار بعامة على بيانات كافية يمكن الاستعانة بها فى تقييم الاختبار.

وكذلك تشتمل على معلومات عن كيفية تطبيق الاختبار وتصحيحه وتفسير درجاته فى ضوء المعايير الخاصة به، وأن تكون الإرشادات والمعلومات الواردة تعبر حقيقة عما يهدف له الاختبار.

كما يجب على مؤلفى الاختبارات أن يعيدوا النظر فى اختباراتهم ومعاييرها كلما دعت الحاجة إلى ذلك حتى لا تتقادم هذه المعايير، مع مراعاة أن الاختبارات تختلف اختلافاً ملحوظاً فى المدة الزمنية اللازمة لمراجعتها.

ومن الضوابط أيضاً فى مجال نشر الاختبارات عدم السماح بالإعلان عن الاختبارات أو نشر أجزاء منها فى الصحف أو المجلات أو الكتب العامة حتى لا يساء فهمها أو استخدامها، وأن يقتصر النشر على الدوريات العلمية المتخصصة بأسلوب موضوعى بعيد عن الانفعال أو الإقناع أو الوصف.

سرية تقارير نتائج الاختبارات :إن من حق المفحوص الاطلاع على تقرير نتائج الاختبارات التى أداها، كما يجب أن يسمح له بالتعليق على محتوى التقرير، وتوضيح أو تعديل بعض المعلومات الخاصة به عند الضرورة. ويهتم المرشدون النفسيون فى الوقت الحاضر بأن يشارك المفحوص مشاركة فاعلة فى عملية تقويمه؛ لذلك يجب أن تقدم إليه نتائج الاختبارات بشكل يسهل عليه فهمه، وأن يبتعد التقرير عن المصطلحات والعبارات الفنية المعقدة.

وإذا كان المفحوص طفلاً يكون من حق والديه الاطلاع على تقرير النتائج، وإن كان هذا يمثل تدخلاً فى خصوصياته (Ruebhausen, 1961)،غير أن التوجيهات التى

اقترحتها مؤسسة ريج (Russell Sage Foundation, 1970) تشير إلى أنه إذا بلغ الطفل من العمر ثمانية عشر عامًا، ولم يتمكن من استكمال دراسته الثانوية، أو تزوج، فإن من حقه أن لا يطلع على مثل هذا التقرير سواه.

غير أن هذا التوجيه أعقبه تحذير حول مدى إمكانية تنفيذ ذلك ومراعاة القوانين المتعلقة بهذا الشأن. ولكن المشكلة هنا ليست مجرد اطلاع القاصر أو والديه على تقرير نتائجه وإنما كيفية إجراء ذلك. فمن الناحية القانونية للوالدين الحق في الحصول على معلومات تتعلق بأبنائهم، وهم عادة يحرصون على ذلك، إلا أنه في بعض الحالات ربما تزداد حدة المشكلات التعليمية والانفعالية للطفل نتيجة عدم توافقه مع والديه. ففي مثل هذه الحالات ينبغي تدخل المرشد النفسي لكي يحصل على المعلومات اللازمة عن الطفل من والديه والعمل على كسب تعاونهم.

أما إذا طُبقت الاختبارات من خلال مؤسسات عمل أو مؤسسات تربوية أو قضائية فإنه ليس من الضروري أخذ موافقة كتابية من الفرد المختبر، وإنما يلزم إشعاره بموعد الاختبار والهدف منه وكيف ستستخدم نتائجه، ومن هم الذين سوف يستخدمونها. والمشكلة المهمة الأخرى تتعلق باستبقاء المعلومات المتعلقة بنتائج الاختبارات النفسية والاحتفاظ بها، فهذه المعلومات تفيد في الدراسات الطولية وتتبع حالة المختبرين. غير أن هذه المعلومات تتقادم بعد مدة زمنية تختلف باختلاف نوعيتها، وبذلك تؤدى إلى استنتاجات غير واقعية، واستخدامها في غير الأغراض التى جمعت من أجلها. لذلك يجب أن تكون هناك ضوابط تنظم عمليات استبقاء هذه المعلومات والحفاظ عليها وعدم النيل منها أو الوصول إليها إلا في الحالات المسموح بها.

ومما يزيد هذه المشكلة تعقيداً التطورات الكبيرة التى حدثت في مجال نظم المعلومات وإمكانية استدعاء المعلومات في سرعة ودقة متناهية، إلا أنه يمكن الإفادة من هذه التطورات في الحفاظ على المعلومات والبيانات المتعلقة بنتائج الاختبارات وابتكار أساليب تضمن سرية هذه النتائج.

خلاصـــة

تناولنا فـى الفصل السابـق بعض أهم الأسس والمعـايير والضوابط الـتى تحكم وتوجِّه عملـيات وممارسات بناء الاختـبارات والمقاييس التربـوية والنفسية، وتطبيقها وتفسير نتائجها. فهناك مشكلات أساسيـة تتعلق بالقياس والتقويم فى الوطن العربى من أهمـها: إسنـاد مهام الـقياس النفسـى والتربـوى إلى غيـر الاختـصاصيين، وخـبرة الاختصاصيين النفسيين المحدودة بنظريات القياس النفسى والتربوى والأسس التى يستند إليهـا، وقلة الاختبارات الـمقننة واختبارات الكفايات، وإسـاءة استخدام الاختـبارات وبخاصـة الاختبـارات التحـصيليـة التى يكتـبها الـمعلمون، وعـدم إيلاء الاختـبارات التشخيصية والبنائية أهمية تذكر.

كما يمكن تصنيـف الأخطاء الناجمة عن إساءة استخدام الاختـبارات والمقاييس النفسية والتربوية فى أخطاء تتعلق بتحيز المسئولين عن البرامج الاختبارية، والثقة التامة فى التنبؤ باستخدام الاختبارات، والاسـتخدام الروتينى لدرجات الاختبارات، وافتراض أن الاختبارات العـقلية تقيس خصائص مـوروثة. أما الأخطاء الناجمة عن تـقديم نتائج الاختبارات إلى المـفحوصين فإنها تتعلـق بمسايرة التوقعات لنتائـج الاختبارات، وتأثر مفهوم الذات لدى الفرد بهذه النتائج.

وهناك العديد من الشروط العلمية لانتقاء الاختبارات والمقاييس التربوية والنفسية من حيث توافـر خصائص الاختبار أو الـمقياس الجيد، ومنـاسبة الاختبار للـمفحوص وللغرض الـذى يستخدم مـن أجله. وكذلك هناك شـروط علمية لـتطبيق الاختـبارات وتصحيحها، وتفـسير درجاتها وتقديم نتائجها. ومن بين ضـوابط وأخلاقيات استخدام هذه الاختبارات والمـقاييس توافر كفايات معينة لدى مستخدمى الاختـبارات، وحقوق المختبرين من حيـث علمهم بالهدف من الاختبار أو المقياس، وفيـم تستخدم نتائجه، وعدم تداول الاختـبارات والمقاييس بين غيـر المتخصصين فى عـلم النفس، وضوابط نشر وتوزيع الاختبارات النفسية، وسرية تقارير نتائج الاختبارات.

الفصل الثالث

أساسيات الإحصاء وتوظيفه في القياس والتقويم التربوى والنفسى

* مقدمة
* أساليب تنظيم وعرض مجموعات من الدرجات
* التمثيل البيانى لتوزيعات الدرجات
* المدرج التكرارى
* المنحنى التكرارى
* شكل الأغصان والأوراق
* استخدام شكل الأغصان والأوراق في المقارنة بين مجموعتين من الدرجات
* وصف توزيعات درجات الاختبارات والمقاييس
* مقاييس النزعة المركزية أو الموضع
* العلاقة بين مقاييس النزعة المركزية أو الموضع
* مقاييس التشتت
* الدرجات المُحوَّلة
* تحديد المركز النسبى للأفراد على المنحنى الاعتدالى
* المئينيات
* الرتب المئينية
* استخدام المنحنى الاعتدالى فى إيجاد المئينيات
* مقاييس العلاقة بين مجموعتين من الدرجات
* معامل ارتباط بيرسون
* استخدام معامل الارتباط فى التنبؤ
* معادلة خط انحدار المتغير (ص) على المتغير (س)
* الخطأ المعيارى للتنبؤ

مقدمة :

ذكرنا فى الفصل الأول أن القياس التربوى والنفسى يهدف للتحديد الكمى للظواهر السلوكية، ويعتمد ذلك على جمع بيانات عن طريق أدوات القياس المتنوعة، وينتج عن تطبيق هذه الأدوات فى معظم الأحيان قيما عددية أو درجات تمثل قياسات لسمات معينة لدى مجموعة من الأفراد.

وهذه القيم أو الدرجات تتطلب تبويبًا وتصنيفًا وتحليلاً لاستخلاص معلومات مفيدة تتعلق بهذه السمات، ويمكن فى ضوء هذه المعلومات تفسير الدرجات المستمدة من هذه الأدوات تفسيراً مستنيراً، كما يمكن في ضوئها التحقق من بعض الخصائص السيكومترية الأساسية لأدوات القياس فى مراحل تصميمها وبنائها والحكم على مدى صلاحيتها فيما تقيسه.

ونظرًا لأن علم الإحصاء يتضمن الطرق والأساليب التى يمكن استخدامها فى معالجة البيانات الكمية والكيفية، فإن معرفة الباحث والممارس لبعض هذه الطرق والأساليب، واكتساب المهارة في توظيفها توظيفًا مناسبًا يمكنه من تفسير الدرجات التى يحصل عليها من أدوات القياس التربوى والنفسى بما يحقق الهدف من استخدام هذه الأدوات.

ونظرًا لتعدد أساليب المعالجات الإحصائية ، فسوف نقتصر فى هذا الفصل على عرض أهم الأساليب الوصفية التى يمكن توظيفها بسهولة فى إعداد أدوات القياس وتفسير نتائجها. وعلى وجه التحديد سوف يشتمل هذا الفصل على الأساليب الإحصائية التى تستخدم فى تنظيم ووصف مجموعات من الدرجات عن طريق مقاييس النزعة المركزية والتشتت والعلاقة، وكذلك الأساليب المختلفة لتحويل الدرجات من صورة إلى أخرى لإجراء مقارنات وظيفية فيما بينها، واشتقاق معايير الاختبارات التربوية والنفسية بما يسمح بتفسير الدرجات المستمدة منها. ونفرد الفصول الأربعة التالية لمناقشة أساليب تحديد الخصائص السيكومترية لأدوات القياس.

أساليب تنظيم وعرض مجموعات من الدرجات :

إن تفرد وتميز الأفراد يُعدان من الخصائص الأساسية للحياة، فالأطفال الذين ينشأون فى الأسرة نفسها أو فى ظروف أسرية متماثلة يختلفون اختلافا واضحا فى كثير من الصفات، ويزداد الاختلاف بينهم فى الاستعدادات والميول والاتجاهات وسمات الشخصية كلما تقدموا فى العمر، ويمكننا ملاحظة هذا الاختلاف بوضوح من خلال تعاملنا مع الأفراد. فبعض الأفراد يتميزون بقدرات ومهارات ربما لا تتوافر فى أقرانهم، مثل القدرة الحسابية أو اللغوية أو الموسيقية أو المهارات اليدوية، وبعضهم يستطيع تكوين صداقات طيبة والبعض الآخر يفضل الوحدة، وهكذا.

وبسبب هذه الاختلافات الواضحة حازت دراسة الفروق الفردية اهتماماً مستمراً من جانب علماء النفس، والدليل على ذلك أن علم النفس الفارق Differential Psychology وهو أحد فروع علم النفس يهتم بالبحث الكمى الموضوعى للفروق الفردية.

ولذلك لا نعجب إذا وجدنا أن معظم المصطلحات المستخدمة فى علم النفس لوصف الأفراد تشير إلى خصائص أو سمات تختلف وتتباين بين الأفراد اختلافا ملحوظًا. وهذه الفروق والاختلافات لها دلالة عملية مهمة ربما يكون أكثرها وضوحًا ما يتطلبه الأداء الجيد فى مختلف المهن والنجاح فى الدراسات الأكاديمية من قدرات ومهارات ربما لا تتوافر بدرجة عالية فى جميع الأفراد.

ولذلك فإن المشكلة الرئيسة التى تهم الباحث والممارس فى المجالات النفسية والتربوية هو كيفية وصف طبيعة ومدى هذه الفروق الفردية وصفًا دقيقًا بالاستعانة بالأساليب الإحصائية التى تساعد فى تنظيم وعرض مجموعات الدرجات المستمدة من أدوات القياس وتمثيلها بيانيا بحيث تفصح عن بعض ما يتميز به الأفراد من خصائص أو سمات. كما أن هذا التنظيم يساعد فى إلقاء الضوء على إجابة بعض التساؤلات المتعلقة بهذه الفروق الفردية.

ولتوضيح ذلك نفترض أن معلم اللغة العربية طبق اختبارًا يقيس فهم المقروء على تلاميذ أحد الصفوف البالغ عددهم ٤٠ تلميذًا ، وكانت الدرجات كالتالى:

٦٧	٧٨	٨٠	١٠٢	١٢٣	١١٤	١٠٧	٦٩	١٣١	٨٠
٩٠	٨٤	٧٩	٨١	٨٩	٩٧	٨٦	٨٢	٨٥	١٢٣
٩١	٨٩	٨١	١١٠	١٢٦	٨٨	١٠٤	٩٦	١٠٣	٨٥
٩٣	٨١	٩٨	٨٢	١٠٧	١٠٠	٧٨	١٠٠	٨٤	٧٣

فبالنظر إلى هذه الدرجات نجد أنها تتراوح بين ٦٧ ، ١٣١ . وعلى الرغم من أن المعلم يستطيع بالفحص العينى لهذه الدرجات ملاحظة أن هناك فروقًا فردية بين التلاميذ، إلا أن استخلاص معلومات دقيقة عن هذه الفروق من مجرد هذا الفحص العينى يتطلب وقتا وجهدًا ملحوظًا.

لذلك ينبغى أن تنظم هذه الدرجات بطريقة تيسر ذلك، فربما يقوم المعلم بترتيب هذه الدرجات ترتيبًا تصاعديًا أو تنازليًا.

غير أن هذا ربما يصعب إجراؤه إذا كان عدد التلاميذ كبيرًا، لذلك يفضل تبويب هذه الدرجات فى فئات متتالية تشتمل كل فئة منها على عدد معين من الدرجات، أى تحديد الفئات ومعرفة عدد التلاميذ الذين يقعون فى كل فئة منها، ووضع فئات الدرجات وتكرار أعداد التلاميذ فى جدول يسمى جدول التوزيع التكرارى Frequency Distribution .

ويمكن اختيار سعة الفئة بحيث ينحصر العدد الكلى للفئات بين ١٠ ، ٢٠ فئة تقريبًا، فإذا قل عدد الفئات بحيث أصبح أربع أو خمس فئات مثلاً تزداد سعة الفئة وبالتالى لا نستفيد شيئًا من عملية التبويب، وإذا زاد عدد الفئات عن ٢٠ فئة تصبح سعة الفئة صغيرة وتضيع معالم التوزيع. وبالطبع يعتمد اختيار سعة الفئة على العدد الكلى للدرجات ومداها والهدف من استخدام التوزيع.

فإذا عدنا إلى مجموعة الدرجات السابقة وأردنا تكوين جدول التوزيع التكرارى، فإن الخطوات تكون كالتالى:

(١) تحديد مدى الدرجات : وهذا المدى يوضح انتشار الدرجات بين القيمتين العليا وهى ١٣١ والدنيا وهى ٦٧. أى أن الدرجات تتوزع فى مدى قدره (١٣١ ـ ٦٧) + ١ = ٦٥ درجة. وأضفنا الواحد الصحيح لأنه ربما يحصل فرد على الدرجة ٦٧ ويحصل آخر على الدرجة ١٣١ ، فعندئذ المدى يزداد بقدر الواحد الصحيح.

(٢) تحديد سعة الفئة : ونعنى بذلك تقسيم المدى إلى فئات متساوية لتبويب الدرجات ، فإذا قسّمنا المدى وقدره ٦٥ إلى ٩ فئات مثلا تكون سعة الفئة مساوية ($\frac{٦٥}{٩}$ = ٧,٢ ونُقرّبها لتصبح ٧).

(٣) تحديد الفئة الدنيا فى التوزيع : وهذا يتطلب البدء بأقل الدرجات بحيث تقبل القسمة على سعة الفئة. ونظرًا لأن أقل الدرجات هى ٦٧ لا تقبل القسمة على سعة الفئة وهى ٧، فإننا نبحث عن الدرجة الأقل مباشرة من ٦٧ وتقبل القسمة على ٧ فنجدها ٦٣. لذلك ينبغى أن تبدأ الفئة الأولى بالدرجة ٣. وتصبح الفئة الدنيا ٦٣ ـ ٦٩،

والفئة الثانية تبدأ بالدرجة التى تلى ٦٩ أى ٧٠، وتكون هذه الفئة ٧٠ ـ ٧٦، وهكذا حتى نصل إلى أعلى الدرجات أو بعدها بقليل.

(٤) تكوين جدول التوزيع التكرارى : فهذا التوزيع عبارة عن مجموعة من الفئات المرتبة ترتيبًا تصاعديًا بحيث تشتمل على جميع الدرجات. والجدول يتكون من ثلاثة أعمدة هى: فئات الدرجات، وعلامات التكرار، والتكرار. ويمكن الحصول على علامات التكرار بأن نمر على الدرجات الأربعين، ونضع علامة (×) فى العمود الثانى أمام الفئة التى تقع فيها كل درجة، ثم نعد العلامات المناظرة لكل فئة وندوّن التكرار فى العمود الثالث، ثم نجمع التكرار الكلى للتأكد من أنه يساوى عدد أفراد المجموعة أى ٤٠.

(٥) إيجاد التكرار المئوى : يُفضل عادة إيجاد نسبة التكرار المناظر لكل فئة إلى التكرار الكلى، وذلك بقسمة كل من التكرارات على العدد الكلى للأفراد وضرب الناتج فى ١٠٠ للحصول على التكرار المئوى.

أى أن : التكرار المئوى المناظر لفئة معينة = $\frac{\text{التكرار المناظر للفئة}}{\text{التكرار الكلى}} \times ١٠٠$

حيث إن هذا يفيد فى مقارنة توزيعات درجات اختبارات مختلفة، وبخاصة إذا كان التكرار الكلى لدرجات كل من الاختبارين مختلفًا، وفيما يلى جدول التوزيع التكرارى للدرجات السابقة:

فئات الدرجات	علامات التكرار	التكرار	التكرار المئوى
٦٣ ـ ٦٩	××	٢	٥,٠
٧٠ ـ ٧٦	×	١	٢,٥
٧٧ ـ ٨٣	××××××××××	١٠	٢٥,٠
٨٤ ـ ٩٠	×××××××××	٩	٢٢,٥
٩١ ـ ٩٧	××××	٤	١٠,٠
٩٨ ـ ١٠٤	××××××	٦	١٥,٠
١٠٥ ـ ١١١	×××	٣	٧,٥
١١٢ ـ ١١٨	×	١	٢,٥
١١٩ ـ ١٢٥	××	٢	٥,٠
١٢٦ ـ ١٣٠	××	٢	٥,٠
		ن = ٤٠	١٠٠,٠

جدول (٣-١) التوزيع التكرارى لمجموعة الدرجات التى عددها ٤٠ درجة

ويلاحظ من جدول (٣ ـ ١) أن الحدود الدنيا للفئات تقبل القسمة على سعة الفئة ٧، وحصلنا على ١٠ فئات وليس ٩ لأننا اعتبرنا سعة الفئة ٧، وليس ٧,٢، كما بدأنا بالدرجة ٦٣ بدلا من ٦٧ لأن الدرجة ٦٣ تقبل القسمة على سعة الفئة ٧.

ومن ميزات هذا الجدول أنه يُمكّن من معرفة توزيع الدرجات وانتشارها بدرجة أفضل من الدرجات غير المبوبة، إذ يستطيع المعلم بمجرد النظر إلى علامات التكرار معرفة أن أكبر عدد من الطلاب حصلوا على درجات تتراوح بين ٧٧ ، ٩٠. ويبين ذلك أيضًا التكرار المئوى للفئتين الثالثة والرابعة وهو ٢٥,٠ + ٢٢,٥ = ٤٧,٥ ٪ من المجموع الكلى للتلاميذ.

بينما حصل عدد أقل من التلاميذ على درجات تقع بين ٦٣، ٧٦ أو بين ١١٩ ، ١٣٢، حيث إن التكرار المئوى فى الحالة الأولى ٧,٥٪ وفى الحالة الثانية ١٠٪، ويمكن أيضًا ملاحظة ذلك بوضوح إذا نظرنا إلى علامات التكرار فى هذه الفئات.

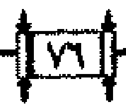
وبالمثل نستطيع تحديد عدد التلاميذ الذين يقعون فى مختلف فئات الدرجات.

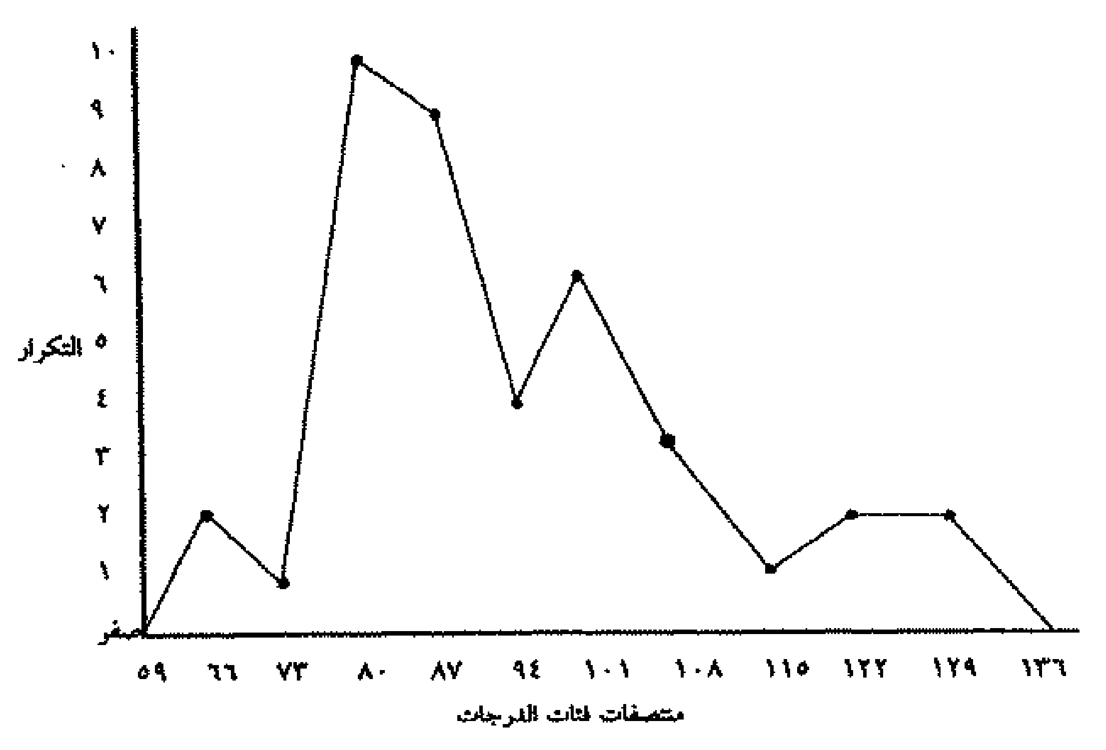
التمثيل البيانى لتوزيعات الدرجات :

على الرغم من أن التوزيع التكرارى لتوزيعات الدرجات يعد أسلوبًا بسيطًا لتنظيم وتلخيص هذه الدرجات، إلا أن التمثيل البيانى للتوزيعات يقدم للطلاب والآباء وغيرهم من المعنيين صورة أكثر وضوحًا وأيسر فهمًا من التوزيع التكرارى للدرجات.

وهناك أنواع متعددة من الأشكال البيانية التى يمكن أن يستخدمها المعلم والباحث والممارس فى وصفه ودراسته للفروق الفردية ، ومن بين هذه الأشكال البيانية شائعة الاستخدام المنحنى التكرارى Frequency Curve ، والمدرج التكرارى Histogram .

المنحنى التكرارى: نظرًا لأن سعة الفئة فى المثال السابق أكبر من الواحد الصحيح، فإن النقاط التى نستخدمها لرسم المنحنى التكرارى تناظر منتصفات الفئات على المحور الأفقى، والتكرارات على المحور الرأسى، ويتكون المنحنى التكرارى من توصيل منتصفات فئات التوزيع بخط منحنى ممهد بقدر الإمكان كما هو موضح بشكل (٣ـ١) التالى:

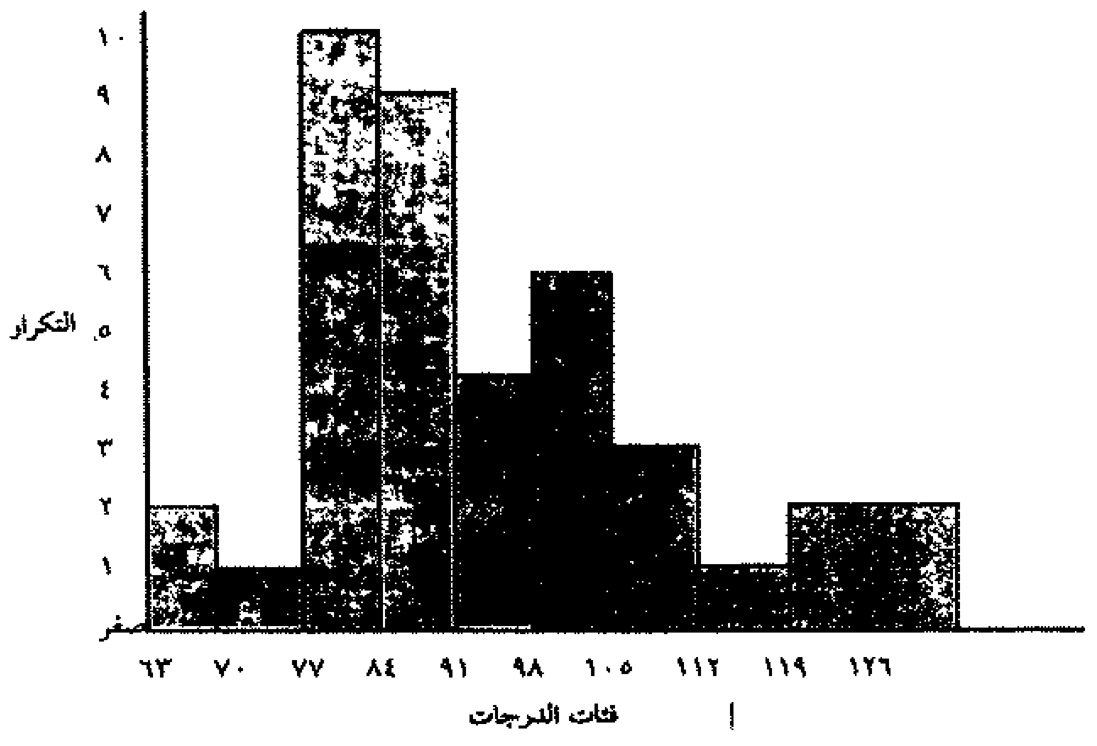




شكل (٣-١) المنحنى التكرارى لدرجات ٤٠ تلميذًا

المدرج التكرارى: عبارة عن مجموعة من المستطيلات المتلاصقة أو المتباعدة تمثل توزيع درجات الاختبارات والمقاييس، فالمستطيلات تكون متلاصقة إذا كان مستوى قياس المتغير الذى تمثله الدرجات فتريا أو نسبيًا، بينما تتباعد المستطيلات عن بعضها البعض إذا كان المتغير كيفيًا، مثل النوع أو الديانة أو المهنة، حيث إن الأقسام التى يصنف فيها كل فرد تكون منفصلة وعندئذ تتباعد المستطيلات.

ففى المثال السابق تمثل فئات الدرجات على المحور الأفقى والتكرار على المحور الرأسى كما فى المنحنى التكرارى، وارتفاع المستطيلات يمثل التكرارات كما هو موضح بشكل (٣ ـ ٢) التالى:



شكل (٣ ـ ٢) المدرج التكرارى لدرجات ٤٠ تلميذاً

ويتضح من الشكلين (٣ ـ ١) ، (٣ ـ ٢) أن كليهما يمثلان مجموعة درجات الاختبار لتلاميذ الفصل، ولكن المنحنى التكرارى يبين انسيابية توزيع الدرجات وشكله العام وحدوده نظراً لأن التوزيع متصل ومستوى القياس فترى.

ومن ميزات المنحنى التكرارى التى ربما لا يتميز بها المدرج التكرارى أنه يمكن استخدامه فى تمثيل توزيعين مختلفين للبنين والبنات مثلا فى اختبار ما، أو توزيع درجات الاختبار نفسه قبل عملية التعليم وبعده، أو قبل أنشطة علاجية وبعدها، أو توزيع درجات اختبارين طبقًا على طلاب فصل معين، وذلك لإجراء مقارنات بين شكل التوزيعين فى كل من هذه الحالات. وعندئذ ربما يتقاطع المنحنيان فى عدة نقاط يمكن تفسيرها، فى حين يصعب إجراء ذلك باستخدام المدرج التكرارى.

ويُفضل تمثيل التكرار المئوى على المحور الرأسى بدلاً من تمثيل التكرار إذا كان عدد أفراد المجموعتين مختلفًا لكى تكون المقارنة دقيقة.

شكل الأغصان والأوراق :

رأينا مما سبق أن التمثيل البيانى لدرجات الاختبارات والمقاييس التربوية والنفسية يوضح الشكل العام لتوزيع الدرجات وييسر استخلاص معلومات مفيدة لمصمم الاختبار ومستخدمه، إلا أن المنحنى التكرارى والمدرج التكرارى ليسا كافيين لاعتمادهما على فئات الدرجات بدلا من الدرجات نفسها، وبذلك نفقد بعض المعلومات المتعلقة بالدرجات، فكل درجة فى التوزيع تمثل أحد أفراد المجموعة التى طبق عليها الاختبار أو المقياس وينبغى الاهتمام بها وتفسيرها.

وشكل الأغصان والأوراق Stem - and - Leaf Display يعد أحد الأساليب الكشفية المستحدثة التى توصل إليها توكى Tukey رائد تحليل البيانات. ويتميز هذا الأسلوب بأنه يعالج بعض أوجه قصور الأساليب التقليدية المستخدمة فى التمثيل البيانى لتوزيعات الدرجات والتى أشرنا إليها. فهذا الأسلوب يجمع بين ترتيب الدرجات ترتيبًا تصاعديًا وتمثيلها تمثيلاً بصريًا، أى يعد بمثابة توزيع تكرارى ومدرج تكرارى للدرجات فى آن واحد، ويحافظ على جميع الدرجات أو القيم الملاحظة مرتبة، وفى الوقت نفسه يساعد فى تنظيم الدرجات بحيث يتضح الشكل العام لتوزيع هذه الدرجات أو القيم.

ولتوضيح كيفية إنشاء شكل الأغصان والأوراق للمثال السابق يمكن اتباع الخطوات التالية :

(١) نُقَسِّم كل درجة من درجات المجموعة إلى جزءين، فإذا كانت الدرجة تتكون من رقمين نفصل رقم خانة العشرات عن رقم خانة الآحاد، فمثلاً الدرجة الأولى ٦٧ تتكون من رقم خانة العشرات ٦ الذى يمثل الغصن Stem، ورقم خانة الآحاد ٧ الذى يمثل الورقة Leaf .

وإذا كانت الدرجة تتكون من ثلاثة أرقام نفصل رقمى خانتى المئات والعشرات عن رقم خانة الآحاد. فمثلا الدرجة ١٢٣ تتكون من رقم خانة المئات ١، ورقم خانة العشرات ٢، ورقم خانة الآحاد ٣.

لذلك نفصل الرقم ١ ، ٢ عن الرقم ٣ وبذلك يكون العدد ١٢ بمثابة الغصن، والرقم ٣ بمثابة الورقة التى تنتمى إلى هذا الغصن . وهكذا فى بقية درجات المجموعة.

(٢) نضع جميع الأغصان فى عمود رأسى إلى اليسار بحيث تكون قيمها مرتبة ترتيبًا تصاعديًا، ونضع الأوراق التى تنتمى إلى كل غصن فى صف أفقى بجوار هذه الأغصان إلى اليمين، ونفصل بين الأغصان والأوراق بخط رأسى. ونظرًا لأن عدد



الدرجات أو القيم يعد أساسيًا، فإنه ينبغى تسجيل جميع الدرجات حتى لو تكرر بعضها لكى لا نفقد أى درجة منها، ويتضح الشكل العام لتوزيع هذه الدرجات. ونظرًا لأن الدرجات تتراوح بين ٦٧ ، ١٣١ فإننا نبدأ عمود الأغصان بالقيمة ٦ وننتهى بالقيمة ١٣. ونظرًا لأننا استخدمنا الدرجات كما هى دون تقريب أو تقليص، فإن الوحدة فى هذه الحالة تكون درجة واحدة.

فإذا اتبعنا هذه الخطوات يمكن أن نحصل على شكل الأغصان والأوراق التالى لمجموعة الدرجات المذكورة فى المثال السابق وعددها ٤٠ درجة :

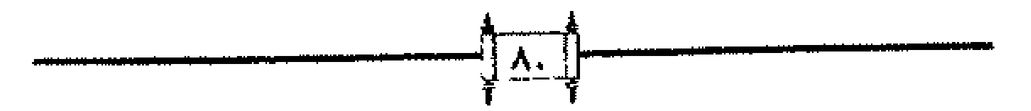
الأغصان	الأوراق
٦	٧٩
٧	٨٩٨٣
٨	٤٩١٠١١٢٩٨٦٢٥٤٠٥
٩	٠١٣٨٧٦
١٠	٢٧٠٧٤٠٣
١١	٠٤
١٢	٣٦٣
١٣	١

شكل (٣ ـ ٣) يوضح الأغصان والأوراق لدرجات ٤٠ تلميذ

ويتضح من شكل (٣ ـ ٣) أن معظم الدرجات تتركز فى الغصن ٨ حيث يحتوى على أكبر عدد من الأوراق (١٥ ورقة) يليه الغصن ١٠ حيث يحتوى على (٧ أوراق) ثم الغصن ٩ الذى يحتوى على (٦ أوراق).

أى أن التوزيع يميل إلى التراكم فى الأغصان ٨، ٩، ١٠. وهذا يعنى أن ٢٨ تلميذًا من بين العدد الكلى للتلاميذ (٤٠ تلميذًا) حصلوا على درجات تتراوح بين ٧٣، ١٠٧. كما يبدو أن التوزيع غير متماثل حول الغصن ٩ الذى ينتمى إليه المتوسط العام للدرجات حيث يوجد التواء إلى حد ما ناحية الأغصان ١١ ، ١٢ ، ١٣ التى ينتمى إليها عدد قليل من الأوراق.

وعلى الرغم من أن شكل الأغصان والأوراق يساعد فى استخلاص أكبر قدر من المعلومات من درجات الاختبارات والمقاييس ويوضح النمط العام لتوزيع الدرجات أو



الشكل العام للتوزيع ، إلا أن هناك بعض الملاحظات التى يجب الإشارة إليها لأهميتها :

(١) **ينبغى عدم تقليل عدد الأغصان كثيراً** : فهذا يؤدى إلى تراكم عدد كبير من الأوراق على كل غصن مما يجعل من الصعب استخلاص معلومات مفيدة تتعلق بتوزيع الدرجات. فمثلاً إذا كانت درجات اختبار ما كالتالى:

٢٧، ١٣، ٢٢، ٢٩، ١٩، ٢٦، ١٤، ٢٥، ٢١، ١٦، ٩، ٢٨، ٧، ٢٦، ١٥.

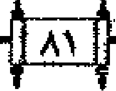
فإننا نلاحظ أن الدرجات تتراوح بين ٧، ٢٩ . فإذا جعلنا الوحدة درجة واحدة فإن عمود الأغصان سوف يشتمل على أرقام خانة العشرات ٢ ، ١ ، صفر فقط، وعندئذ تتراكم الأوراق المناظرة لأوراق خانة الآحاد على هذه الأغصان الثلاثة مما يجعل الشكل قليل الفائدة.

ولذلك يُفضّل فى هذه الحالة توسعة شكل الأغصان والأوراق وذلك بتقسيم الأغصان. فمثلا يمكن تقسيم كل غصن إلى غصنين أحدهما للأوراق المناظرة للدرجات ٥ وما فوقها ، والآخر للدرجات ٤ وأقل. وبذلك يصبح شكل الأغصان والأوراق مكونًا من ستة أغصان بدلاً من ثلاثة ، كما هو موضح بشكل (٣ ـ ٤) التالى :

الأوراق	الأغصان
	صفر
٩ ٧	صفر*
٣ ٤	١
٩ ٦ ٥	١*
٢ ١	٢
٧ ٩ ٦ ٥ ٨ ٦	٢*

* تشير إلى الأوراق المناظرة للدرجات ٥ وما فوقها

شكل (٣ ـ ٤) يوضح الأغصان والأوراق لدرجات ١٥ تلميذاً بعد تقسيم كل غصن إلى غصنين



ويمكن أيضًا تقسيم كل غصن إلى عدد أكبر من الأغصان وليكن أربعة أقسام، وبذلك يصبح الشكل مكونًا من تسعة أغصان كما يوضحه شكل (٣ ـ ٥) التالى :

الأغصان	الأوراق
صفر ●	
صفر *	
صفر ■	٧ ٩
١ ●	
١ *	٥ ٤ ٣
١ ■	٦ ٩
٢ ●	١ ٢
٢ *	٥
٢ ■	٦ ٨ ٦ ٩ ٧

● تشير إلى الأوراق صفر ، ١ ، ٢ .

* تشير إلى الأوراق ٣ ، ٤ ، ٥ .

■ تشير إلى الأوراق ٦ وما فوقها . .

شكل (٣ ـ ٥) يوضح الأغصان والأوراق لدرجات ١٥ تلميذًا بعد تقسيم كل غصن إلى أربعة أغصان

وعمومًا، فإنه يمكن توسعة شكل الأغصان والأوراق أو تضييقه إما بتقسيم كل صف إلى صفين أو أكثر أو بضم صفين متجاورين أو أكثر، وهذا يعتمد على طبيعة البيانات والمعلومات المراد استخدامها من الشكل.

(٢) ينبغى عدم الإكثار من الأغصان : فكثرة الأغصان يؤدى إلى خلوها من الأوراق أو قلتها، ويحدث هذا غالبًا عندما تكون القيم مشتملة على أربعة أرقام أو أكثر، ولذلك يفضل فى مثل هذه الحالات تقليص الشكل وذلك باختيار وحدة كبيرة ربما تكون ١٠ درجات أو ١٠٠ درجة أو ربما أكثر بحسب قيم الدرجات، وكذلك تقريب هذه القيم تقريبًا مناسبًا.

ولتوضيح ذلك نفترض أن القيم كانت كالتالى:

١٠٢١٢	١٢٥٢٧ ،	١١١٢١ ،	٩٣٥٢ ،	١٢١٣٨ ،
١١١٣٧	١٤٢٧٧ ،	١٣٢٧٦ ،	١٠٩٤٤ ،	١١٦٨٧ ،
١٦١٢٢	١١٢٢٧ ،	١٠٨٩١ ،	١٥٢٤١ ،	١٢٢٣١ ،

فعندئذ لا تصلح أن تكون الوحدة درجة واحدة، وإنما يفضل هنا أن تكون الوحدة ١٠٠ مثلا وتقريب القيم إلى أقرب مائة، فالقيمة ١٠٢١٢ تصبح ١٠٢٠٠، وكذلك القيمة ١٢٥٢٧ تصبح ١٢٥٠٠، وهكذا.

كما يُفضل أن يُعبر عن قيم الأغصان بالألف. فمثلاً القيمة ١٠٢٠٠ تقسم إلى جزءين ١٠، ٢ حيث يوضع العدد ١٠ فى عمود الأغصان والرقم ٢ فى صف الأوراق التى تنتمى إلى هذا الغصن، وبذلك يكون شكل الأغصان والأوراق لهذه المجموعة من الدرجات كالتالى (شكل ٣ ـ ٦):

الأغصان	الأوراق
٩	٣
١٠	٨ ٩
١١	١ ٢ ١ ٦
١٢	٢ ١ ٥
١٣	٢
١٤	٢
١٥	٢
١٦	١

الوحدة ١٠٠ درجة

الغصن ١٠٠٠ درجة

شكل (٣ ـ ٦) يوضح الأغصان والأوراق لمجموعة من الدرجات باستخدام وحدة ١٠٠ درجة

وبالطبع يمكن أيضًا تقسيم كل غصن إلى عدد أكبر من الأغصان إذا تطلب التوزيع التوسعة، أو ضم صفوف أوراق متجاورة إذا كانت كثير من الأغصان بدون أوراق أو أوراقها قليلة.

ففي الشكل (٣ ـ ٦) يلاحظ انتشار الأوراق رغم قلة عددها؛ وذلك لأن العدد الكلي للدرجات قليل، ولذلك ربما تحتاج الصفوف المتجاورة قليلة الأوراق إلى الضم، مع مراعاة أن تظل الفروق الفردية بين الدرجات واضحة بحيث يمكن أن نستخلص منها أكبر قدر من المعلومات المفيدة عن التوزيع.

استخدام شكل الأغصان والأوراق في المقارنة بين مجموعتين من الدرجات:

سبق أن أشرنا إلى كيفية المقارنة بين الشكل العام لكل من مجموعتين من الدرجات باستخدام المنحنى التكراري، وسوف نوضح كيفية إجراء ذلك باستخدام شكل الأغصان والأوراق. وقد لاحظنا أننا نستخدم النظام العددي في عمود الأغصان بدلا من سعة الفئة في جدول التوزيع التكراري، ولعل هذا ييسر إنشاء شكل الأغصان والأوراق في وقت قصير، ولكن هذا يؤدي إلى مقارنة غير مناسبة بين مجموعات مختلفة من الدرجات. فإذا كان لدينا مجموعتين من الدرجات وأمكن تمثيلهما بشكلين مختلفين من أشكال الأغصان والأوراق ولكن بعمود أغصان مشترك، فإنه ينبغي أن تكون الوحدة المستخدمة لقيم أو درجات كل من المجموعتين متساوية، فلا يجوز أن تكون الوحدة لإحدى المجموعتين درجة واحدة، وتكون الوحدة ١٠ درجات للمجموعة الأخرى وبخاصة إذا كان التوزيعان لهما نفس النمط أو الشكل العام. إذ يصعب معرفة الارتفاع النسبي لكل من التوزيعين في حالة اختلاف الوحدة المستخدمة في عمود الأغصان مما يؤدي إلى مقارنات غير دقيقة. ولعل هذا أيضًا ما جعلنا نشير إلى استخدام التكرار المئوي بدلاً من التكرار عند المقارنة بين مجموعتين من الدرجات باستخدام المنحنى التكراري كما سبق أن أوضحنا.

ولمعرفة كيفية إجراء مقارنة بين مجموعتين من الدرجات في اختبار ما باستخدام شكل الأغصان والأوراق، نفترض أنه أمكن تمثيل مجموعتين من الدرجات بشكل الأغصان والأوراق التالي:

أوراق المجموعة الثانية	الأغصان	أوراق المجموعة الأولى
	٤	١ ٥
٣ ٢	٥	٦ ٧ ١ ٣
٨ ٧	٦	٨ ٢ ٠ ٤ ٥ ٢
	٧	٩ ٣ ٢ ٤ ٠
٢ ٤ ٥	٨	٧ ٣ ١ ٢ ٥ ٦ ٨
٣ ٨	٩	٣ ٢ ٠ ٥ ٥ ٤ ١ ٢ ٣
٧	١٠	١ ٢ ٧ ٣
٨ ٥ ٢ ١ ٠ ٩ ٧	١١	٤ ٥
٣ ٥ ٣ ٨ ٤	١٢	
٤ ٧ ٢ ١	١٣	٢

الوحدة درجة واحدة لكل من المجموعتين

شكل (٣ ـ ٧) يوضح الأغصان والأوراق لمجموعتين من الدرجات

وبالنظر إلى شكل (٣ ـ ٧) يتضح أنه على الرغم من أن درجات المجموعة الأولى تتراوح بين ٤١، ١٣٢ والثانية بين ٥٢، ١٣١ ، أي أن انتشار كل من المجموعتين متقارب، إلا أننا نلاحظ أن ذيل توزيع المجموعة الأولى من الدرجات يتجه بدرجة أكبر نحو الدرجات الدنيا بينما يتجه ذيل توزيع المجموعة الثانية نحو الدرجات العليا. كما أن درجات المجموعة الأولى تتركز بين ٦٠، ٩٥ (٦٧,٥٪ من الدرجات) بينما تتركز درجات المجموعة الثانية بين ١١٠، ١٢٨ (٦٤٪ من الدرجات) مما يدل على اختلاف الشكل العام لكل من التوزيعين، فأحدهما يميل إلى الالتواء تجاه اليمين والآخر تجاه اليسار، وهذا يعني أن مستوى أداء أفراد المجموعة الثانية أفضل من مستوى أداء أفراد المجموعة الأولى.

وبالطبع كلما كان شكل الأغصان والأوراق لكل من المجموعتين أكثر تماثلاً، فإن ذلك يجعل التفسير أيسر. فإذا كان التوزيعان مختلفين اختلافا ملحوظًا فإنه يمكن إجراء تحويلات معينة على الدرجات، مثل التحويل اللوغاريتمي لمعالجة اختلاف الشكل العام للتوزيعين وجعلهما أكثر تماثلاً. وسوف نتناول بعض أنواع أخرى من التحويلات في هذا الفصل تمكننا من المقارنة بين توزيعات الدرجات بدقة أكثر . ومن

الجدير بالذكر أن هناك برامج للحاسوب يمكن باستخدامها الحصول على أشكال الأغصان والأوراق وإجراء التحويلات المختلفة على الدرجات مثل حزمة برامج MINITAB التى يمكن إدارتها ببيئة النوافذ Windows .

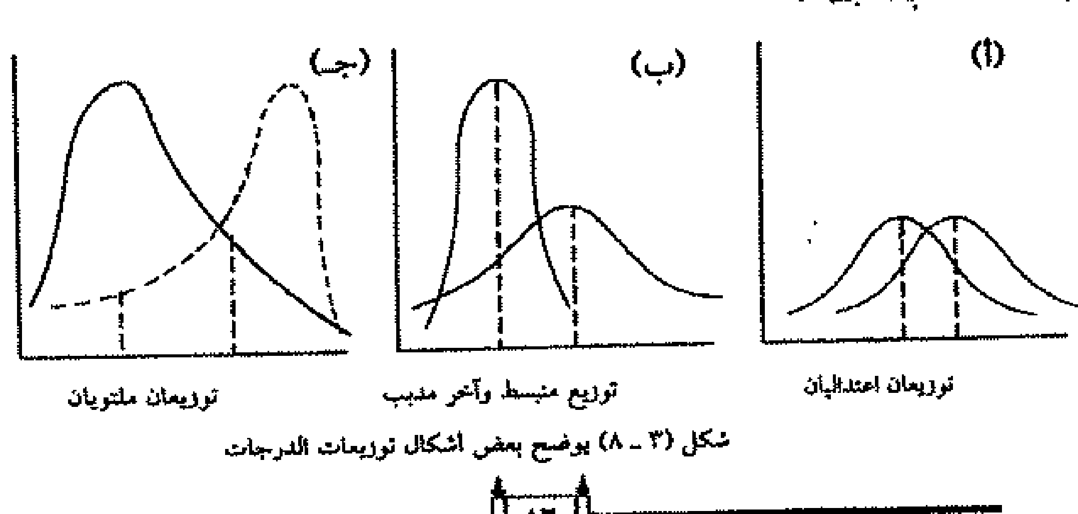
وصف توزيعات درجات الاختبارات والمقاييس :

عرضنا فى الجزء السابق كيفية تنظيم درجات الاختبارات والمقاييس التربوية والنفسية بطرق متعددة كخطوة تمهيدية لوصف توزيعات هذه الدرجات، وإلقاء الضوء على الفروق بين الأفراد فى السمة التى يقيسها الاختبار أو المقياس.

ويعتمد توزيع هذه الفروق الفردية على طبيعة السمة المقاسة فى المجتمع وعينة الأفراد ومحتوى الاختبار أو المقياس الذى يستخدم فى قياس هذه السمة. غير أننا إذا بحثنا ظاهرة نفسية معينة مثل الذكاء أو التوافق أو القلق لدى عينة من طلاب المدارس الثانوية مثلا، فإننا نجد أن درجات مجموعة كبيرة من هؤلاء الطلاب الذين يتم انتقاؤهم عشوائيًا من مجتمع طلاب المدارس الثانوية بعامة يتخذ شكلا يشبه الجرس المقلوب ويسمى المنحنى الاعتدالى Normal Curve حيث نجد عددًا قليلا نسبيًا من ذوى الذكاء المنخفض، وعددًا قليلاً نسبيًا من ذوى الذكاء المرتفع، بينما نجد أن غالبية الطلاب ذكاؤهم متوسط، أى يقل التكرار كلما اتجهنا نحو طرفى المنحنى بعيدًا عن المتوسط.

وهذه الخاصة تنطبق فى كثير من الأحيان على معظم الظواهر النفسية والتربوية، أى الميل إلى التمركز أو النزعة المركزية Central Tendency مع مراعاة الشروط التى أشرنا إليها.

وتختلف التوزيعات التكرارية للدرجات بالطبع باختلاف خصائص التوزيع، فمنها ما هو منبسط ومنها ما هو مدبب ومنها ما هو ملتو التواء موجبًا أو سالبًا، فهذه التوزيعات تختلف فى النزعة المركزية وكذلك فى التشتت كما هو مبين بالأشكال التالية (٣ ـ ٨ أ ، ب، جـ) .



شكل (٣ ـ ٨) يوضح بعض أشكال توزيعات الدرجات

مقاييس النزعة المركزية أو الموضع :

إن الوصف الكمى للسمات الإنسانية يتطلب مقاييس إحصائية مناسبة، إذ لا يكفى أن نفحص توزيع درجات الاختبارات والمقاييس التربوية والنفسية من حيث النمط أو الشكل العام وتراكم الدرجات أو تطرفها، وإنما ينبغى أن نحصل على قيم تمثل النزعة إلى التراكم حول درجة معينة. ومن أهم مقاييس النزعة المركزية أو الموضع Measures of Location: المتوسط Mean، والوسيط Median ، والمنوال Mode . ولكل من هذه المقاييس مزاياها وعيوبها واستخداماتها. ويعتمد اختيار المقياس المناسب على الشكل العام لتوزيع الدرجات، ومستوى قياس المتغير الذى يمثل سمة معينة، والتساؤلات التى نود الإجابة عنها. لذلك سوف نوضح هذه المقاييس الثلاثة فيما يلى:

المتوسط الحسابى: يعد من أهم مقاييس النزعة المركزية أو الموضع حيث إنه شائع الاستخدام فى تفسير درجات الاختبارات، فكثيراً ما نقول أن درجة فرد ما فى أحد الاختبارات تفوق المتوسط أو تقل عن المتوسط. كما يعد أكثر هذه المقاييس استخدامًا فى دراسة الفروق الفردية والتعبير عنها بقيمة واحدة متوسطة.

ويمكن ببساطة الحصول على قيمة المتوسط الحسابى لمجموعة من الدرجات بقسمة مجموع هذه الدرجات على عددها، وعادة يرمز للمتوسط الحسابى بالحرف سَ.

أى أن : سَ = $\frac{\text{مجـ س}}{\text{ن}}$ .... (٣ ـ ١) .

حيث ترمز (مجـ س) إلى مجموع الدرجات .

، وترمز (ن) إلى عدد الدرجات .

فإذا كان لدينا الدرجات ٢، ٤، ٦، ٨، ١٠ مثلاً،

فإن المتوسط الحسابى سَ = $\frac{٢+٤+٦+٨+١٠}{٥}$ = $\frac{٣٠}{٥}$ = ٦.

وعلى الرغم من أن الصيغة (٣ ـ ١) يمكن استخدامها إذا كان عدد الدرجات كبيراً، إلا أنه يفضل تبويب الدرجات فى هذه الحالة فى جدول يتكون من عمودين أحدهما يحتوى على الدرجات والآخر يحتوى على التكرارات المناظرة للدرجات.

فمثلاً إذا كان لدينا الدرجات :

٢، ٤، ١٠، ٦، ١٠، ٦، ٤، ٨، ١٠، ٦، ٨، ١٠

فإننا نكوّن الجدول التالى:

الدرجـــــــة (س)	تكرار الأفراد (ت)
٢	١
٤	٢
٦	٣
٨	٢
١٠	٤
المجموع	١٢

ولإيجاد قيمة المتوسط الحسابى نضرب كل درجة (س) فى التكرار المناظر لها (ت) ونجمع الناتج، ثم نقسم هذا الناتج على عدد الأفراد.

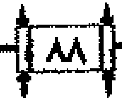
أى أن : $\bar{س} = \frac{مجـ (س \times ت)}{مجـ ت}$ ...... (٣ ـ ٢)

$= \frac{(٢\times١) + (٤\times٢) + (٦\times٣) + (٨\times٢) + (١٠\times٤)}{١٢}$

$= \frac{٢ + ٨ + ١٨ + ١٦ + ٤٠}{١٢} = \frac{٨٤}{١٢} = ٧.$

ونظراً لأن دراسة الفـروق الفردية دراسة علميـة والتوصل إلى نتائج كمـية يمكن الاستناد إليها فى تفسير هذه الفروق يتطلب تطبيق الاختبارات أو المقاييس المناسبة على عدد كبير نسبيًا من الأفراد، وإيجاد المتوسط الحسابى لهذا العدد الكبير من الدرجات باستخدام إحدى الطريقـتين السابقتين يتطلب جهداً كبيراً، لذلـك ينبغى استخدام برامج الحاسوب فى إجراء ذلك.

الوسيط : أحيانًا يكون تـوزيع درجات اختبار ما ملتويـا كما فى شكل (٣ ـ ٩) التالـى، كأن يكون الاخـتبار صـعبًا أو سـهلاً لكثـير من الأفراد الـذين يجرى علـيهم الاختبار. ونظراً لأن المتوسط الحسابى يتأثر كـثيراً بالقيم المتطرفة فى التوزيع أى القيم



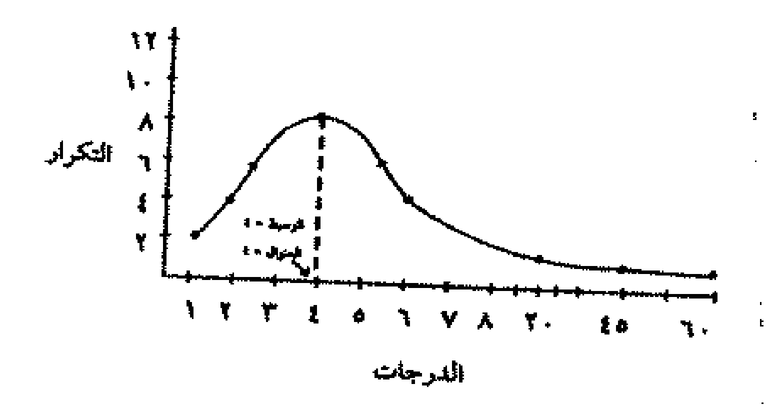
الكبيرة أو الصغيرة نسبيا، فإنه لا يصلح استخدامه فى هذه الحالة لأنه سيؤدى إلى قيم غير ممثلة للنزعة المركزية، ولذلك يفضل استخدام الوسيط.

ولتوضيح ذلك نفترض أن لدينا توزيع الدرجات التالية الموضح بجدول (٣ ـ ٢) التالى:

الدرجات	١	٢	٣	٤	٥	٦	٢٠	٤٥	٦٠
التكرار	٢	٣	٥	٨	٧	٤	٢	١	١

جدول (٣ ـ ٢) توزيع تكرارى لمجموعة من الدرجات

ويمكن تمثيل هذا التوزيع بالمنحنى التكرارى الموضح بشكل (٣ـ٩) التالى:



شكل (٣ ـ ٩) يوضح المنحنى التكرارى لمجموعة الدرجات الموضحة بالجدول (٣ـ٢)

ويتضح من شكل (٣ ـ ٩) الذى يمثل المنحنى التكرارى لتوزيع الدرجات المذكورة فى جدول (٣ ـ ٢) أن هناك ذيلا للمنحنى تجاه الدرجات المرتفعة ٢٠، ٤٥، ٦٠، أى أن المنحنى ملتو التواء موجبًا. فقيمة المتوسط الحسابى فى هذه الحالة باستخدام المعادلة (٣ ـ ٢) تساوى ٧,٩ تقريبًا، وهى قيمة لا تعبر بدقة عن النزعة المركزية للتوزيع؛ لأنها تأثرت تأثرًا واضحًا بالقيم المتطرفة تجاه ذيل المنحنى، ولذلك يفضل استخدام الوسيط فى هذه الحالة. وهذا يتطلب ترتيب الدرجات ترتيبًا تصاعديًا أو تنازليا، وإيجاد القيمة التى تتوسط الدرجات. فالوسيط إذن هو القيمة التى تقسم التوزيع إلى قسمين متساويين.

فإذا كان عدد الدرجات فرديًا تكون هناك قيمة واحدة، أما إذا كان عددها زوجيًا تكون هناك قيمتان نوجد متوسطهما.

ونظرًا لأن تكرار الدرجات الموضحة فى جدول (٣ ـ ٢) ٣٣ أى فرديًا، فإن هناك قيمة واحدة للوسيط. وفيما يلى الترتيب التصاعدى لهذه الدرجات:

١١ ٢٢٢ ٣٣٣٣٣ ٤٤٤٤٤٤[٤]٤ ٥٥٥٥٥٥٥ ٦٦٦٦ ٢٠ ٢٠ ٤٥ ٦٠

ترتيب الوسيط = $\frac{٣٣ + ١}{٢}$ = ١٧.

قيمة الوسيط = ٤ ، وذلك لأن الدرجة ٤ التى بداخل المربع تقسم التوزيع إلى قسمين متساويين.

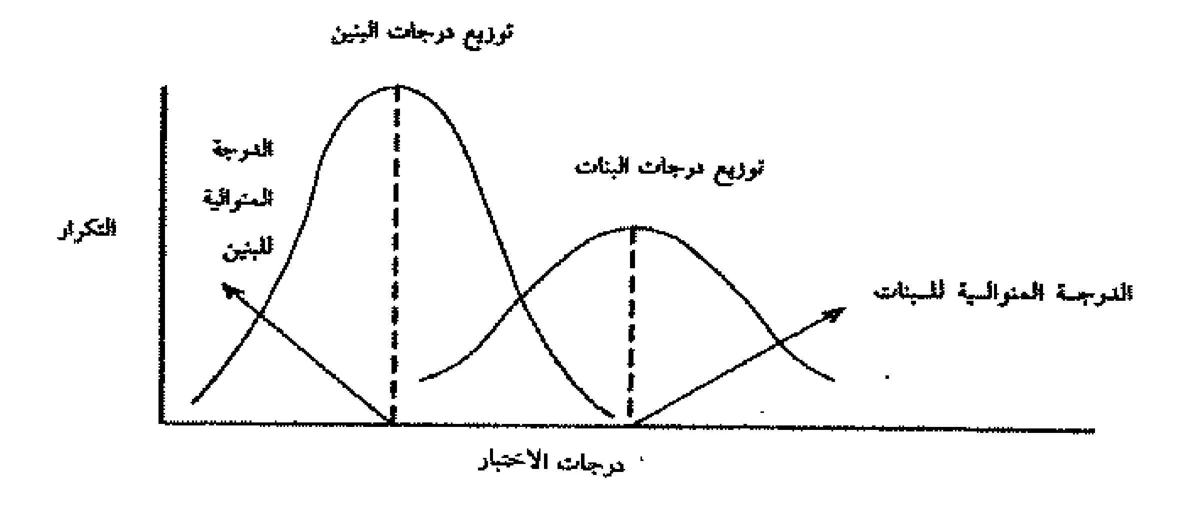
ويلاحظ أن قيمة المتوسط أكبر من قيمة الوسيط بدرجة واضحة وذلك لتأثر المتوسط بالقيم ٢٠، ٤٥ ، ٦٠ بينما لم يتأثر الوسيط.

لذلك فإن الوسيط فى هذه الحالة أكثر تمثيلاً ودقة للنزعة المركزية من المتوسط نظرًا لالتواء التوزيع. ويلاحظ أن قيمة الوسيط اعتمدت على ترتيب الدرجات فقط دون إجراء أى عمليات حسابية، وذلك لأنه يتطلب أن يكون القياس من المستوى الرتبى على الأقل.

المنوال : إذا قلنا أن ذكاء فرد ما متوسط، فإن هذا ربما يعنى أن الدرجة التى حصل عليها فى اختبار الذكاء هى الدرجة الأكثر شيوعًا بين أقرانه، وعندئذ تكون درجته هى الدرجة المنوالية.

فالمنوال إذن هو القيمة الأكثر شيوعًا أو تكرارًا فى توزيع الدرجات، فقيمة منوال الدرجات المذكورة فى جدول (٣ ـ ٢) تساوى ٤ نظرًا لأن هذه الدرجة تكررت أكثر من غيرها من الدرجات. وبالنظر إلى التوزيع الملتوى الموضح بشكل (٣ ـ ٩) نلاحظ أن الدرجة ٤ تناظرقمة التوزيع التى يحددها الخط الرأسى المتقطع، وتمثل الدرجة المنوالية لأنها تناظر أكبر تكرار وهو ٨.

وأحيانًا يكون للتوزيع قيمتين أى منوالين مما يدل على تركيز الدرجات عند أكثر من موضع. فإذا طبقنا اختبارًا على مجموعتين إحداهما من البنين والأخرى من البنات، فإننا ربما نجد أن قمة المنحنى التكرارى لتوزيع درجات البنين تختلف عن قمة المنحنى التكرارى لتوزيع درجات البنات كما فى شكل (٣ ـ ١٠) التالى:



شكل (٣ ـ ١٠) يوضح منحنيان تكراريان لدرجات مجموعتين مختلفتين

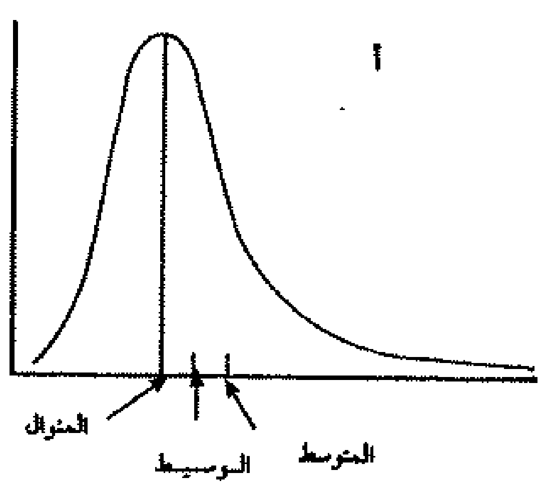
ويتضح من شكل (٣ ـ ١٠) أنه على الرغم من أن التكرار المناظر للدرجة المنوالية للبنين أكبر من التكرار المناظر للدرجة المنوالية للبنات، إلا أن قيمة الدرجة المنوالية للبنات أكبر من قيمتها للبنين، فالدرجة المنوالية للبنات تفوق الدرجة المنوالية للبنين على الخط الأفقى الذى يمثل درجات الاختبار.

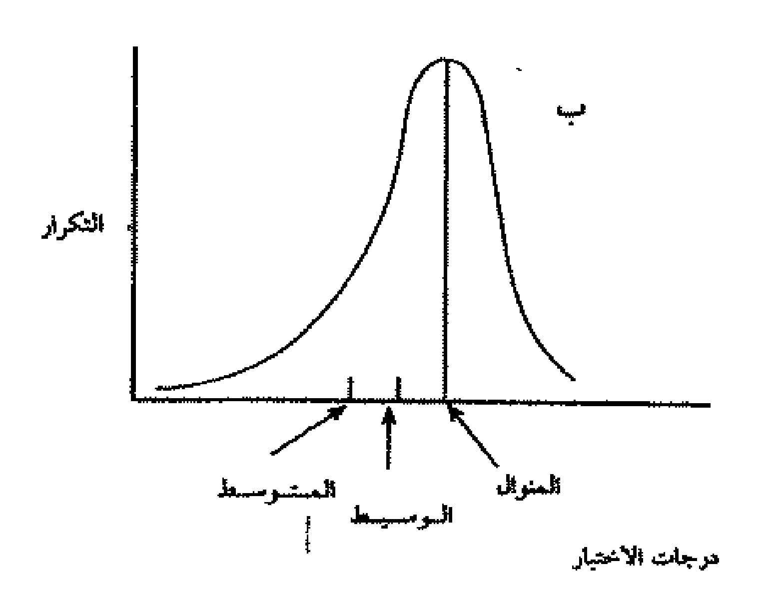
وعلى الرغم من أن المنوال يعد أحد مقاييس الموضع، إلا أن قيمه سريعة التذبذب، حيث إن إضافة أو حذف أى درجة من درجات التوزيع تؤدى إلى تغيير قيمته جوهريًا، ولكنه يناسب المتغيرات التصنيفية أى مستوى قياسها يكون اسميًا.

فمثلاً الطالب الذي يُنتخب رائدًا للفصل هو ذلك الطالب الذى يحظى بأكبر عدد من الأصوات من أقرانه، كما أنه يعد أبسط مقاييس الموضع من حيث عملياته الحسابية.

العلاقة بين مقاييس النزعة المركزية أو الموضع :

ذكرنا فيما سبق أن التوزيعات التكرارية يمكن أن تكون اعتدالية أى متماثلة أو ملتوية التواء موجبًا أو سالبًا، فإذا كان التوزيع اعتداليًا أو متماثلاً وأحادى المنوال، فإن قيم المتوسط والوسيط والمنوال تكون متساوية، أما إذا كان التوزيع أحادى المنوال وملتويا التواء موجبًا أو سالبًا فإن ذيل التوزيع يتجه نحو الدرجات العليا أو الدرجات الدنيا على الترتيب كما هو موضح بالشكلين (٣ ـ ١١أ) ، (٣ ـ ١١ب) على الترتيب:





شكل (٣ ـ ١١) يوضح العلاقة بين مقاييس الموضع إذا كان التوزيع ملتويا التواء موجبا أو سالبا

فإذا نظرنا إلى شكل (٣ ـ ١١أ) نلاحظ أن قيمة المتوسط تأثرت بالقيم العليا المتطرفة على الخط الأفقي الممثل لدرجات الاختبار، بينما نجد قيمة الوسيط تقسم التوزيع إلى قسمين متساويين. وقيمة المنوال هي الدرجة المناظرة لأكبر تكرار الذي يمثله الخط الرأسي من قمة المنحنى. ففي هذه الحالة تكون قيمة المتوسط أكبر من قيمة الوسيط، وقيمة الوسيط أكبر من قيمة المنوال.

وإذا نظرنا أيضًا إلى شكل (٣ ـ ١١ ب) نلاحظ أن العكس صحيح، حيث إن قيمة المتوسط أصبحت أصغر من قيمة الوسيط لتأثره بالقيم الدنيا المتطرفة على خط الدرجات، وقيمة الوسيط أصغر من قيمة المنوال.

لذلك ينبغي مراعاة ما يلي عند استخدام مقاييس النزعة المركزية، أو الموضع لوصف توزيعات درجات الاختبارات والمقاييس التربوية والنفسية وصفا كميًا :

(١) إذا كان مستوى قياس المتغير أو السمة فتريا أو نسبيًا يفضل استخدام المتوسط الحسابي، وإذا كان مستواه رتبيا يستخدم الوسيط، أما إذا كان المتغير اسميا أي تصنيفيا فالمنوال يكون مناسبًا في هذه الحالة.

(٢) إذا كان توزيع الدرجات ملتويًا التواءً ملحوظًا، أي يشتمل على درجات

متطرفة، فإنه يفضل استخدام الوسيط كمقياس للموضع،حيث إن الوسيط لا يتأثر كثيراً بهذه الدرجات، وإنما يتأثر فقط بعدد الدرجات، ويسهل عندئذ إيجاد قيمة الوسيط بترتيب الدرجات ترتيبًا تنازليًا أو تصاعديًا واختيار الدرجة الوسطى.

(٣) بعض مقاييس التشتت التي سنتناولها بعد قليل مثل الانحراف المعياري والتباين يعتمدان في حسابهما على المتوسط. لذلك إذا تطلب الأمر وصف تشتت توزيع الدرجات كميًا، فإنه يفضل استخدام المتوسط كمقياس للموضع.

وعمومًا، فإنه يصعب وصف توزيع الدرجات بقيمة واحدة حيث إننا نفقد بذلك كثيرًا من المعلومات المتعلقة بالتوزيع، لذلك فإنه يفضل إيجاد قيم المتوسط والوسيط والمنوال، ويكون لدى مستخدم الاختبار معلومات تفصيلية عن توزيع درجاته، حيث إن كلا من هذه المقاييس الثلاثة تقدم معلومات مفيدة مختلفة.

مقاييس التشتت :

إن مقاييس النزعة المركزية أو الموضع ليست كافية للمقارنة بين توزيعات درجات الاختبارات المختلفة أو بين توزيعات درجات اختبار ما طبق على مجموعات مختلفة من الأفراد، فلو أن متوسط درجات تلاميذ أحد فصول الصف السادس في اختبار الحساب مثلا ٣٥، ومتوسط درجات تلاميذ فصل آخر في الاختبار نفسه ٣٥، فإننا لا نستطيع استنتاج أن مستوى تحصيل كل من الفصلين متكافئ، إذ ربما تكون درجات كل تلميذ في الفصل الأول ٣٥ ، بينما حصل بعض تلاميذ الفصل الثاني على درجات أعلى بكثير من ٣٥ ، وقلة منهم حصلت على درجات منخفضة، ومع هذا فإن المتوسط في كل من الحالتين يمكن أن تكون قيمته ٣٥.

فإذا نظرنا إلى مجموعتي الدرجات التاليتين :

٣٥ ، ٣٥ ، ٣٥ ، ٣٥ ، ٣٥

٤٥ ، ٥٥ ، ٦٠ ، ٨ ، ٧

نجد أن قيمة المتوسط في كل من الحالتين ٣٥، ومع هذا فإن درجات المجموعة الثانية أكثر تشتتًا من درجات المجموعة الأولى، بل إن المجموعة الأولى تشتتها صفرًا.

لذلك ليس من الصواب أن نقارن بين المجموعتين على أساس المتوسط بمفرده، وإنما يلزم أيضًا أخذ قيمة تشتت كل من المجموعتين بعين الاعتبار لكي تكون المقارنة دقيقة وواقعية.

أى أننا يجب أن نهتم عند دراسة الفروق الفردية فى إحدى السمات بمتوسطات مجموعات الأفراد فى هذه السمة وكذلك تشتت درجات كل مجموعة حول متوسطها.

وهناك مقاييس متعددة للتشتت، ولكننا سوف نقتصر على المقاييس التى يمكن الإفادة منها فى تفسير درجات الاختبارات والمقارنة بينها. وهذه المقاييس هى: المدى، والانحراف المعيارى، والتباين.

المدى : يعد المدى Range أبسط مقاييس التشتت فى حسابه، إذ يعتمد على الفروق بين أكبر درجة وأصغر درجة فى مجموعة الدرجات، فمثلاً :

مدى الدرجات ١٥، ١٢، ١١، ٩ ، ٦ يساوى ١٥ - ٦ = ٩

ومدى الدرجات ٤٨، ٤٨، ٤٨، ٤٨، ٤٨، ٣٩ يساوى ٤٨ - ٣٩ = ٩

أى أن مدى درجات كل من المجموعتين متساو.

ولكن إذا تأملنا مجموعتى الدرجات نجد أن المجموعة الثانية أكثر تجانسًا أى أقل تشتتًا من المجموعة الأولى، وربما يرجع هذا التناقض إلى أن المدى يعتمد على قيمتين فقط، فهو يتأثر تأثرًا بالغًا بإضافة درجة أو قيمة واحدة إلى الدرجات مما يجعله مضللا كما رأينا ، فالمدى لا يأخذ القيم الوسطى فى الحسبان.

لذلك لابد من الاعتماد فى حساب التشتت على مقاييس أكثر دقة أهمها الانحراف المعيارى Standard Deviation ، والتباين Variance .

الانحراف المعيارى : يعد أكثر مقاييس التشتت شيوعًا واستخدامًا فى دراسة الفروق الفردية، بل ويعد أكثرها دقة، فهو يعتمد على جميع درجات التوزيع وليس على درجتين فقط كما فى المدى.

ونحصل على قيمة الانحراف المعيارى لمجموعة من الدرجات بأن نوجد انحراف كل درجة عن المتوسط، ونربّع كل من هذه الانحرافات، ونجمع الناتج، ونقسمه على عدد الدرجات، ثم نستخرج الجذر التربيعى لخارج القسمة.

أى أن الانحراف المعيارى ع = $\sqrt{\frac{\text{مجـ}(\text{س} - \bar{\text{س}})^2}{\text{ن}}} = \sqrt{\frac{\text{مجـ ح}^2}{\text{ن}}}$

فإذا كانت الدرجات ٨، ٨، ٩، ٩، ٩، ٩، ١٠، ١٠ ، فإنه يفضل تكوين جدول كالتالى (٣ ـ ٣) :

الدرجات (س)	س - $\bar{س}$ = ح	(س - $\bar{س}$)^{٢} = ح^{٢}
٨	٨ - ٩ = -١	١
٨	٨ - ٩ = -١	١
٩	٩ - ٩ = صفر	صفر
٩	٩ - ٩ = صفر	صفر
٩	٩ - ٩ = صفر	صفر
٩	٩ - ٩ = صفر	صفر
١٠	١٠ - ٩ = +١	١
١٠	١٠ - ٩ = +١	١
المجموع مجـ س = ٧٢ المتوسط $\bar{س}$ = ٩	مجـ (س - $\bar{س}$) = صفر	مجـ (س - $\bar{س}$)^{٢} = ٤

جدول (٣ ـ ٣) يوضح خطوات حساب قيمة الانحراف المعيارى لمجموعة من الدرجات

يتضح من جدول (٣ ـ ٣) أن مجـ (س - $\bar{س}$)^{٢} = ٤ .

$$\text{فعندئذ تكون ع} = \sqrt{\frac{\text{مجـ}(س - \bar{س})^{٢}}{ن}} = \sqrt{\frac{\text{مجـ ح}^{٢}}{ن}}$$

$$= \sqrt{\frac{٤}{٨}} = \sqrt{٠,٥} = ٠,٧١ \text{ تقريباً}$$

ويمكن تبسيط العمليات الحسابية إذا استخدمنا الصيغة (٣ـ٤) التالية لإيجاد قيمة الانحراف المعيارى :

$$ع = \sqrt{\frac{\text{مجـ}(س^{٢}) - \frac{(\text{مجـ س})^{٢}}{ن}}{ن}} \quad ......(٣ - ٤)$$

وهذه الصيغة لا تتطلب إيجاد انحرافات الدرجات عن المتوسط، وإنما تتطلب فقط إيجاد مجموع مربعات الدرجات (مجـ س^{٢}).

ففى المثال السابق مجـ س ^{٢} = ٦٥٢ .

$$، ع = \sqrt{\frac{\frac{٦٥٢ - (٧٢)^{٢}}{٨}}{٨}}$$

$$= \sqrt{\frac{٦٥٢ - ٦٤٨}{٨}}$$

$$= \sqrt{\frac{٤}{٨}} = \sqrt{٠,٥} = ٠,٧١ \text{ تقريبًا}$$

وهى نفس القيمة السابقة.

وكلما اقتربت قيمة الانحراف المعيارى من الصفر دل ذلك على قلة التشتت. ونظرًا لأن القيمة ٠,٧١ قريبة من الصفر، فإن معنى ذلك أن مجموعة الدرجات متجانسة، ويمكن ملاحظة ذلك إذا نظرنا إلى الدرجات الثمانية.

التباين : هو مربع الانحراف المعيارى أى :

$$ع^{٢} = \frac{مجـ (س - \bar{س})^{٢}}{ن} = \frac{مجـ ح^{٢}}{ن} \quad ...(٣ـ٥)$$

فتباين الدرجات فى المثال السابق = (٠,٧١)^{٢} = ٠,٥٠ وهى قيمة قريبة أيضًا من الصفر مما يدل على تجانس الدرجات وقلة تشتتها.

طريقة تقريبية لإيجاد قيمة الانحراف المعيارى :

يمكن إيجاد قيمة تقريبية للانحراف المعيارى والتباين ببساطة وسرعة إذا كان عدد الدرجات ٢٠ أو أكثر، مثل درجات الاختبارات التحصيلية الصفية التى يختبر المعلم بها تلاميذه فى فصل معين؛ وذلك لأن القيم الناتجة تقترب من القيم الفعلية إذا لم يقل

عدد الدرجـات عن ذلك، وتصبح القيم مساوية تقريبًا للقيم الفعلية إذا كـان عدد الدرجات ٣٦ أو أكثر. وقد اقترح لاثروب Lathrop الصيغة التالية لإجراء ذلك :

$$ع = \frac{\text{مجموع السدس الأعلى من الدرجات - مجموع السدس الأدنى}}{\text{نصف (عدد الدرجات - ١)}} \quad (٦\text{-}٢)$$

ويمكن تطبيق هذه الصيغة على مجموعة الدرجات المذكورة في جدول (٣ - ٣) بأن نرتب الدرجات ترتيبًا تنازليًا كالتالي:

١٠ ، ١٠ ، ٩ ، ٩ ، ٩ ، ٩ ، ٨ ، ٨

ثم نأخذ $\frac{١}{٦}$ الدرجات أي $\frac{١}{٦} \times ٨ = ١,٣$ درجة .

ولكننا سنقربها إلى درجة واحدة لكل من السدس الأعلى والسدس الأدنى.

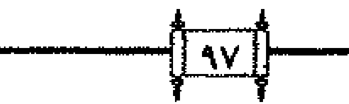
$$ع = \frac{١٠ - ٨}{\frac{١}{٢}(٨ - ١)} = \frac{٢}{٣,٥} = ٠,٥٧ \text{ تقريبًا .}$$

وإذا قارنّا هذه القيمة التقريبية بالقيمة التي حصلنا عليها باستخدام الصيغة (٣ -٣) نجد أن الفرق ٠,١٤ فقط ، ويرجع هذا الفرق إلى قلة عدد الدرجات، فكلما زاد العدد قل هذا الفرق بحيث يحتمل ألا يتعدى ٠,٠١ أو ٠,٠٢ على الأكثر.

استخدامات الانحراف المعياري :

(١) يمكن استخدام الانحراف المعياري والمتوسط معًا للمقارنة بين مجموعتين أو أكثر من الدرجات، نظرًا لأن المتوسط بمفرده لا يصلح لذلك كما سبق أن أشرنا.

ولكن يشترط أن تكون الاختبارات المستمدة منها الدرجات لـها نفس وحدات القياس، فالانحراف المعياري يختلف بقدر اختلاف وحدات القياس. لذلك إذا طبق اختباران أحدهما في الاستعداد اللفظي والآخر في القراءة على طلاب فصل واحد، ووجد أن قيمة الانحراف المعياري لدرجات الاختبار الأول أكبر من قيمته للاختبار الثاني، فلا يجوز القول بأن تشتت درجات الطلاب في الاختبار الأول أكبر من تشتت درجاتهم في الاختبار الثاني. فهذا الاستنتاج يتطلب التأكد من أن وحدات ميزان قياس الاستعداد الدراسي هي نفسها وحدات ميزان قياس التحصيل في القراءة. ولذلك يفضل أن تجرى المقارنات إذا طبق اختبار واحد على مجموعتين أو أكثر من الأفراد وذلك لوصف التجانس النسبي للمجموعتين في السمة التي يقيسها الاختبار.



(٢) يمكن أيضًا استخدام الانحراف المعيارى فى تحديد الأوزان النسبية للدرجات المركبة من مجموع درجات اختبارين أو أكثر: فبعض اختبارات الذكاء ومقاييس الشخصية تتكون من عدة اختبارات أو مقاييس فرعية مستقلة. والدرجة الكلية عبارة عن مجموع درجات هذه الاختبارات الفرعية، وكذلك درجة التحصيل العام للطالب تكون عادة مركبة من مجموع درجاته فى عدة اختبارات شهرية مستقلة، وفى مثل هذه الحالات لا يجوز جمع الدرجات الفرعية جمعًا بسيطًا، وإنما يكون الجمع موزونًا. ونقصد بالجمع الموزون مراعاة الأهمية النسبية لكل درجة فرعية من حيث إسهامها فى الدرجة المركبة. فأوزان الدرجات الفرعية تعتمد على الانحراف المعيارى لهذه الدرجات بالنسبة لغيرها من الدرجات التى تشكل الدرجة المركبة، فكلما زاد تشتت مجموعة من الدرجات زاد وزنها النسبى. لذلك ينبغى أن نعطى وزنًا لكل درجة فرعية يعادل قيمة الانحراف المعيارى لدرجات الاختبار الفرعى المناظر. ويجوز أيضًا أن نجعل الوزن مساويًا قيمة الانحراف المعيارى مضروبًا فى أى عامل نراه مناسبًا لأهمية الاختبار الفرعى . ولتوضيح ذلك نفترض أن لدينا اختبارًا ما يشتمل على ثلاثة اختبارات فرعية مستقلة أى غير مرتبطة، وأردنا الحصول على درجة مركبة من درجات الاختبارات الفرعية الثلاثة ،ووجدنا أن الانحراف المعيارى لهذه الاختبارات ٧، ٦، ١٠ على التوالى، ولكننا رأينا أنه نظرًا لأهمية الاختبار الثانى فإننا سوف نضاعف قيمة الانحراف المعيارى لهذا الاختبار، وفى الوقت نفسه رأينا تقليل وزن الاختبار الثالث لقلة أهميته، وذلك بقسمة قيمة انحرافه المعيارى على ٢ ، وبذلك تصبح أوزان الدرجات الفرعية ٧، ١٢، ٥. وفى هذه الحالة تكون الدرجة المركبة مساوية لمجموع الدرجات التى يحصل عليها الفرد فى الاختبارات الفرعية بعد ضرب كل درجة منها فى الوزن المناظر لها، أى أن الدرجة المركبة تكون مساوية لمجموع الدرجات الموزونة بقيم الانحراف المعيارى أو مضاعفاته أو أجزائه.

(٣) يمكن كذلك استخدام الانحراف المعيارى فى إجراء تحويلات خطية على الدرجات:

وتفيد هذه التحويلات فى تحديد مركز الفرد الذى حصل على درجة معينة فى الاختبار بالنسبة لأقرانه الذين طبق عليهم الاختبار نفسه. كما تفيد فى إجراء مقارنات أكثر دقة وواقعية بين مجموعات من الدرجات المستمدة من اختبارات تختلف فى متوسطاتها وانحرافاتها المعيارية.

وبذلك تساعد هذه التحويلات فى تكوين جداول معايير الاختبارات والمقاييس التربوية والنفسية. ونظرًا لأهمية ذلك فإننا سوف نناقشه فيما يلى بمزيد من التفصيل.

الدرجات المُحولة :

تستخدم هذه الدرجات فى تحديد المركز النسبى للفرد فى توزيع ما، بحيث يمكن وصف أداؤه بالنسبة لأقرانه فى اختبار يقيس سمة معينة.

ولعل هذا مرتبط ارتباطًا وثيقًا بما سبق أن ذكرناه فى الفصل الأول عن الاختبارات مرجعية الجماعة أو المعيار Norm - Referenced Tests ، فهذه الاختبارات تهدف لتقدير أداء الفرد بالنسبة لأداء الأفراد الآخرين فى السمة أو القدرة التى يقيسها الاختبار، ويكون الاهتمام هنا منصبًا على ترتيب الأفراد بعضهم بالنسبة للبعض الآخر من حيث درجاتهم فى الاختبار، ويكون للدرجة تفسيراً فقط فى ضوء متوسط أداء الجماعة أو ما يسمى بمعيار الجماعة Group Norm ،وتصبح الدرجة لا معنى لها ويصعب تفسيرها إذا أغفلنا ردها إلى هذا المعيار.

وتوجد أنواع مختلفة من الدرجات المحولة لعل أكثرها استخدامًا فى القياس التربوى والنفسى الدرجات المعيارية Standard Scores، والدرجات التائية T-Scores، ويعتمد هذان النوعان من الدرجات على التحويلات الخطية Linear Transformations لدرجات الاختبار باستخدام المتوسطات الحسابية والانحرافات المعيارية معًا.

الدرجات المعيارية

ذكرنا فيما سبق أن إجراء مقارنات بين الدرجات التى يحصل عليها فرد فى اختبارات مختلفة يتطلب اشتراكها فى وحدة القياس نفسها. ويمكن إجراء ذلك بتحديد عدد وحدات الانحراف المعيارية التى تبتعد فيها درجة الفرد فى كل من هذه الاختبارات عن متوسطات درجاتها، وتسمى هذه الوحدات «الدرجات المعيارية Scores Z» .

فإذا افترضنا أن الدرجات التى حصل عليها طالب فى أربعة اختبارات تختلف فى متوسطاتها، وانحرافاتها المعيارية كانت كالتالى (جدول ٣ ـ ٤) :

الاختبار	درجة الطالب س	المتوسط (س̄)	الانحراف المعيارى (ع)
الأول	٩	١٠	١
الثانى	١٠	١٣	٣
الثالث	٧	٧	٢
الرابع	١٠	١٣	٢

جدول (٣ ـ ٤) يوضح درجات طالب فى أربعة اختبارات ومتوسطات درجات الاختبارات وانحرافاتها المعيارية

ويتضح من جدول (٣ ـ ٤) أن درجة الطالب في كل من الاختبارات الأول والثاني والرابع تقل عن متوسط درجات أقرانه بقدر ١ ، ٣ ، ٣ وحدة انحراف معياري، بينما تقع درجته في الاختبار الثالث عند المتوسط.

وفي ضوء هذه الانحرافات عن المتوسطات يبدو أن أداء الطالب في الاختبارين الثاني والرابع منخفضاً بالقدر نفسه مع أن تشتت درجات الاختبار الثاني أكبر، حيث إن الانحراف المعياري لدرجات هذا الاختبار ٣.

فكلما تراكمت الدرجات حول المتوسط يزداد المركز النسبي للدرجات المرتفعة، وعلى العكس من ذلك كلما زاد التشتت يقل المركز النسبي لهذه الدرجات بين المجموعة.

لذلك لا ينبغي الاقتصار على الانحرافات عن المتوسط، وإنما يجب أخذ قيم الانحراف المعياري أيضاً بعين الاعتبار عند مقارنة درجات الطالب في الاختبارات الأربعة، وذلك للتغلب على مشكلة اختلاف التوزيعات في المتوسطات والانحرافات المعيارية، وهذا يتطلب قسمة انحرافات الدرجات عن متوسطاتها على قيم الانحراف المعياري المناظرة لها، ويسمى الناتج عندئذ الدرجات المعيارية (د) أي أن :

$$\text{د} = \frac{\text{س} - \bar{\text{س}}}{\text{ع}} \quad \ldots\ldots\ldots \quad (\text{٣ ـ ٦})$$

ومن الجدير بالذكر أن متوسط الدرجات المعيارية = صفر، وانحرافها المعياري = الواحد الصحيح.

وبذلك تكون الدرجات المعيارية المناظرة لدرجات الطالب في الاختبارات الأربعة على الترتيب كالتالي:

$$\text{د}_{\text{١}} = \frac{\text{٩} - \text{١٠}}{\text{١}} = \text{١-}$$

$$\text{، د}_{\text{٢}} = \frac{\text{١٠} - \text{١٣}}{\text{٣}} = \text{١-}$$

$$\text{، د}_{\text{٣}} = \frac{\text{٧} - \text{٧}}{\text{٣}} = \text{صفر}$$

$$\text{، د}_{\text{٤}} = \frac{\text{١٠} - \text{١٣}}{\text{٢}} = \text{١,٥-}$$

ويمكن الآن المقارنة بين درجات الطالب فى الاختبارات الأربعة بالنسبة لأقرانه، فمركزه النسبى بين أقرانه فى الاختبار الثالث أفضل من مركزه النسبى فى بقية الاختبارات على الرغم من أنه حصل على الدرجة ٧ فى هذا الاختبار وهى أقل من درجاته فى الاختبارات الأخرى.

ولذلك فإن الدرجة ٧ التى تعد درجة خام لا تصلح للمقارنة إلا إذا تم تحويلها إلى درجة معيارية لكى يتحدد موقعها بالنسبة لمتوسط الدرجات.

ونظراً لأن الدرجات المعيارية هى درجات مجردة ليس لها وحدة خاصة ، فإنه يمكن جمع الدرجات المعيارية للطالب فى الاختبارات الأربعة للحصول على درجة معيارية مركبة. كما يمكن تعيين أوزان مختلفة لهذه الدرجات قبل حساب الدرجة المركبة. فإذا أردنا أن يسهم الاختبار الأول بنصف ما يسهم به الاختبارين الثانى والثالث، وأن يسهم الاختبار الرابع بقدر مرة ونصف، فإن الدرجة المركبة تكون عندئذ كالتالى:

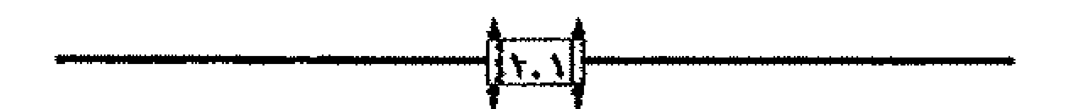
الدرجة المركبة = ٠,٥٠ د١ + د٢ + د٣ + ١,٥ د٤

بعض عيوب الدرجات المعيارية :

على الرغم من أهمية الدرجات المعيارية فى إجراء المقارنات بين درجات الاختبارات المختلفة، إلا أن هناك بعض المشكلات العملية المتعلقة بها. فالمعلم ومستخدم الاختبار ينبغى عليه تفسير هذه الدرجات للطالب أو المفحوص فى ضوء المتوسط والانحراف المعيارى، كما أن الدرجات المعيارية تكون أحياناً سالبة أو كسرية مما يزيد من صعوبة تفسيرها.

فإذا نظرنا إلى المثال السابق نجد أن درجة الطالب فى كل من الاختبارين الأول والثانى تبعد بقدر انحراف معيارى واحد سالب عن متوسط أقرانه، ودرجته فى الاختبار الرابع تبعد بقدر ١,٥ انحراف معيارى، أى أنه يقل عن متوسط أقرانه فى الاختبارات الثلاثة، فهنا يصعب تبرير القيم السالبة والكسرية للطالب أو المفحوص أو ولى الأمر.

ويمكن التغلب على هذه المشكلات بإجراء تحويلات معينة على الدرجات المعيارية والحصول على درجات معيارية معدلة تهدف للتخلص من الإشارات السالبة والكسور العشرية التى تنتج عن تطبيق الصيغة (٣ ـ ٦) وبذلك يسهل تفسير هذه الدرجات وفهمها.



وتوجد عدة تحويلات يمكن إجراؤها على الدرجة المعيارية، غير أننا سوف نقتصر على التحويلات التي تستخدم بكثرة في إعداد معايير الاختبارات التربوية والنفسية، مثل الدرجة التائية T -Scores، والمئينيات Percentiles .

الدرجات التائية:

تعد الدرجات التائية (ت)درجات معيارية معدّلة، وتنتج عن إجراء تحويل خطي للدرجات المعيارية (د). ونقصد بالتحويل الخطي أن نضرب كل قيمة من قيم الدرجات المعيارية (د) في مقدار ثابت وجمع مقدار ثابت آخر على ناتج الجمع. ولعل أفضل مثال يوضح ذلك هو تحويل درجات الحرارة من ميزان مئوي إلى ميزان فهرنهيتي تحويلاً خطيًا، فضرب جميع القيم في مقدار ثابت يغير من سعة فترات ميزان القياس، وجمع مقدار ثابت يغير من وضع الصفر الاعتباري على الميزان بقدر مساو لهذا المقدار الثابت، ولكن عمليتي الضرب والجمع لا تغيران العلاقات القائمة بين وحدات الميزان.

ولذلك فإن الصيغة العامة للتحويلات الخطية للدرجات المعيارية إلى درجات معدلة تكون كالتالي:

الدرجة المعيارية المعدلة = أ + ب × الدرجة المعيارية . . . . . . (٣ ـ ٧)

حيث (أ) ، (ب) مقداران ثابتان.

وعلى الرغم من أن قيمة كل من (أ) ، (ب) اختيارية، إلا أن ماكول McCall جعل المتوسط ٥٠ بدلا من الصفر، والانحراف المعياري ١٠ بدلاً من الواحد الصحيح لكي نستطيع التخلص من الإشارات السالبة والقيم الكسرية للدرجات المعيارية.

والصيغة التالية تستخدم في إجراء هذا التحويل باستخدام الصيغة العامة (٣ـ٧) :

ت = ٥٠ + ١٠ × د . . . .(٣ ـ ٨)

وباختصار فإنه يمكن تحويل الدرجة المعيارية (د) إلى درجة تائية (ت) بأن نضرب الدرجة المعيارية في ١٠ ونجمع ٥٠ على حاصل الضرب.

ومن الجدير بالذكر أنه يمكن اختيار أي قيم أخرى لكل من المتوسط والانحراف المعياري تختلف عن ٥٠ ، ١٠ بحسب مقدار الدرجات.

فإذا كانت الدرجات الخام للاختبار قيمها كبيرة بدرجة واضحة فإنه يمكن تصميم نظام مختلف لتحويل هذه الدرجات إلى درجات معيارية معدلة ، إذ ربما نجعل المتوسط ٥٠٠ والانحراف المعياري ١٠٠ كما في اختبارات الاستعداد الدراسي، أو نجعل المتوسط ١٠٠ والانحراف المعياري ١٦ كما في اختبارات الذكاء وهكذا، وذلك بتطبيق الصيغة العامة (٣ ـ ٧).

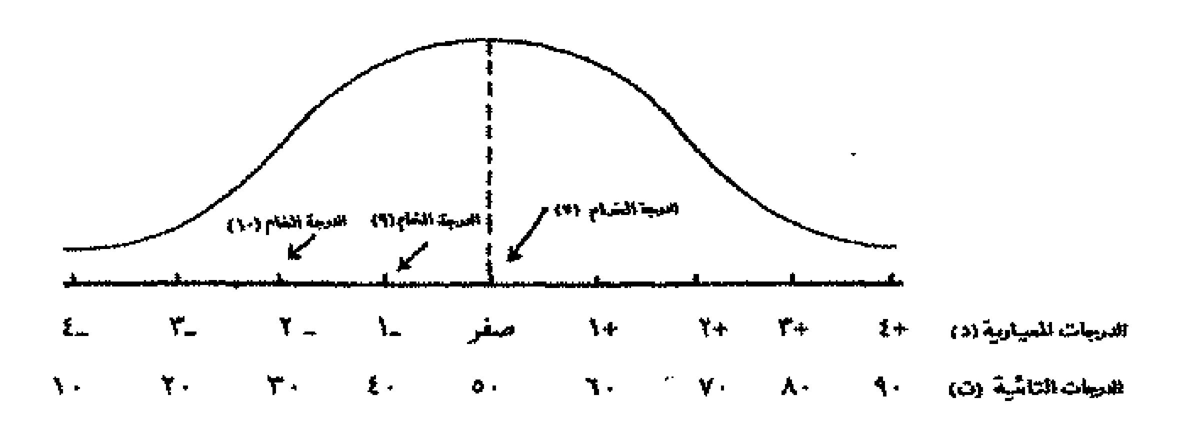
ولتوضيح كيفية تطبيق صيغة الدرجة التائية (ت)، نعود إلى جدول (٣ ـ ٤)، فنجد أن الدرجات التائية للطالب المناظرة للدرجات المعيارية للاختبارات الأربعة كالتالي (جدول ٣ ـ ٥):

الاختبار	الدرجة الخام (س)	الدرجة المعيارية د = (س − سَ) / ع	الدرجة التائية ت = ٥٠ + ١٠ × د
الأول	٩	− ١	٤٠
الثانى	١٠	− ١	٤٠
الثالث	٧	صفر	٥٠
الرابع	١٠	− ١,٥	٣٥

جدول (٣ ـ ٥) يوضح الدرجات المعيارية والتائية المناظرة للدرجات الخام فى أربعة اختبارات

ويتضح من جدول (٣ ـ ٥) أن الدرجة الخام ٩ تناظر الدرجة المعيارية −١ ، وهذه بدورها تناظر الدرجة التائية ٤٠، وهذا يعنى أن الدرجة الخام ٩ تنحرف بقدر انحراف معيارى واحد عن المتوسط، وهكذا فى بقية الدرجات.

ويلاحظ أن الدرجات التائية قد تخلصت من الإشارات السالبة والقيم الكسرية فى الدرجات المعيارية (د) ، وبذلك يصبح تفسيرها معقولاً . ونظراً لأن الدرجات المعيارية (د) تمثل توزيعًا له مَعْلمين ثابتين هما : المتوسط = صفر ، والانحراف المعيارى = ١ ، فإن الدرجات التائية تمثل التوزيع نفسه بمَعْلمين ثابتين أيضًا هما المتوسط = ٥٠، والانحراف المعيارى = ١٠. والدرجات المعيارية التى تفوق المتوسط (صفر) تنحرف عنه بقدر + ١ ، +٢ ، +٣ ، . . . . انحراف معيارى، بينما تنحرف الدرجات المعيارية التى تقل عن المتوسط بقدر −١ ، −٢ ، −٣ . . . . ، انحراف معيارى. أما الدرجات التائية فتنحرف عن المتوسط (٥٠) بقدر ١٠ وحدات معيارية إلى اليمين وإلى اليسار. ويمكن توضيح ذلك بشكل (٣ ـ ١٢) التالى:



شكل (٣ ـ ١٢) يوضح الدرجات المعيارية والدرجات التائية

ويتضح من شكل (٣ ـ ١٢) أن الدرجات المعيارية تتراوح عمليًا بين -٣ ، +٣ وحدة انحراف معياري، بينما تتراوح الدرجات التائية عمليا أيضًا بين ٢٠ ، ٨٠ وجميعها قيم صحيحة موجبة.

كما يتضح من الشكل المركز النسبي للطالب في الاختبارات الأربعة حيث إنه لم يتخط المتوسط في أي من الاختبارات، بل إن درجاته في ثلاثة اختبارات تقل عن المتوسط حيث تناظر الدرجات التائية ٤٠، ٤٠، ٣٠، مما يدل على انخفاض أداؤه فيها بالنسبة لأقرانه.

والخلاصة أن الدرجات المعيارية والدرجات المحولة تحدد المركز النسبي للفرد في اختبار يقيس سمة معينة بوحدات انحرافات معيارية عن متوسط أقرانه الذين طبق عليهم الاختبار نفسه.

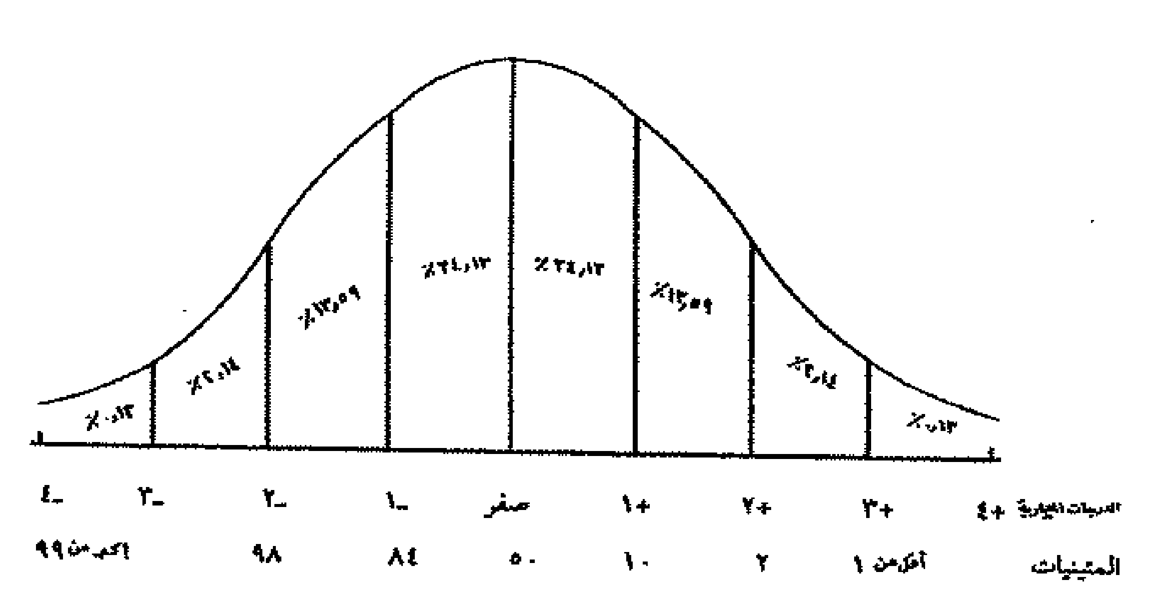
تحديد المركز النسبي للأفراد على المنحنى الاعتدالي :

اتضح لنا مما سبق أن الاختبارات مرجعية الجماعة أو المعيار تهدف لتحديد المركز النسبي للفرد بين أقرانه في السمة التي يقيسها الاختبار.

لذلك يعمل مصممو هذا النوع من الاختبارات على إبراز الفروق الفردية والكشف عنها بحيث تتوزع الدرجات توزيعًا اعتداليًا. فالتوزيع الاعتدالي يمكن تمثيله بيانيًا بمنحنى يعرف بالمنحنى الاعتدالي Normal Curve أو منحنى جاوس، ويتميز هذا المنحنى بأنه يشبه الجرس المقلوب، وهو منحنى متصل ومتماثل حول القيمة المنوالية التي تناظر أقصى ارتفاع للمنحنى، ويمتد من كلتا الجهتين إلى اللانهاية، أي يقترب

تدريجيًا من المحور الأفقى، ولكنه لا يمسه مهما مددناه من كلتا الجهتين. والمساحة الكلية تحت المنحنى، أى المساحة المحصورة بين المنحنى والمحور الأفقى تساوى الواحد الصحيح.

وتتميز جميع المنحنيات الاعتدالية بأن نسبة الأفراد التى تنحصر درجاتهم بين درجتين معياريتين مختلفتين ثابتة، إذ يمكن باستخدام مبادئ علم التكامل فى الرياضيات التوصل إلى أن حوالى ٣٤,١٣٪ من درجات التوزيع الاعتدالى تقع بين المتوسط والدرجة المعيارية +١ ، ١٣,٥٩ ٪ من الدرجات تقع بين الدرجتين المعياريتين +١، +٢، وحوالى ٢,١٤٪ من الدرجات تقع بين الدرجتين +٢، +٣. ونظراً لتماثل المنحنى فإن نفس النسب المئوية من الدرجات تقع بين الدرجات المعيارية السالبة. ويوضح شكل (٣ ـ ١٣) التالى هذه النسب على المنحنى الاعتدالى:



شكل (٣ ـ ١٣) يوضح النسب المئوية من الحالات تحت المنحنى

الاعتدالى المحصورة بين الدرجات المعيارية المختلفة

ومن مزايا هذا المنحنى الاعتدالى المعيارى أنه يمكن استخدامه فى تحويل الدرجات المعيارية إلى مئينيات Percentiles ، أى نسبة مئوية من عدد الأفراد الذين تقل درجاتهم عن درجة معيارية معينة كما سيتضح بعد قليل.

المئينيات :

تعد المئينيات Percentiles من أكثر الطرق استخدامًا في تفسير درجات الاختبارات والمقاييس التربوية والنفسية، فالمئينيات يمكن تفسيرها بسهولة للطالب والمفحوص والمعنيين بنتائج أدوات القياس. كما يمكن تحويل الدرجات المعيارية والدرجات التائية وغيرها من الدرجات المعيارية المعدلة إلى مئينيات باستخدام المنحنى الاعتدالي المعياري الموضح بشكل (٣ ـ ١٣) .

فالمئيني هو نقطة على توزيع الدرجات تقع عندها أو أقل منها نسبة معينة من الأفراد الذين طبق عليهم اختبار معين.

فالوسيط مثلاً يناظر المئيني ٥٠ لأنه يمثل نقطة تقسيم التوزيع إلى نصفين متساويين كما ذكرنا من قبل. فإذا نظرنا إلى شكل (٣ ـ ١٣) نجد أن الفرد الذي يحصل على درجة معيارية +٢ يتفوق على حوالي ٩٨٪ من أقرانه وبذلك فإن درجته المعيارية تناظر المئيني ٩٨، والفرد الذي يحصل على درجة معيارية - ١ يتفوق على حوالي ١٦ ٪ من أقرانه، ودرجته المعيارية تناظر المئيني ١٦ ، وهكذا.

ويمكن أيضًا تقسيم توزيع الدرجات إلى عشرة أقسام متساوية يسمى كل منها إعشاري Decile ، فالإعشاري التاسع مثلاً يناظر المئيني ٩٠. كما يمكن تقسيم التوزيع إلى أربعة أقسام متساوية يناظر كل منها إرباعي Quartile ، فالإرباعي الأول (ر١) يناظر المئيني ٢٥، والإرباعي الثاني (ر٢) يناظر الوسيط أي المئيني ٥٠، والإرباعي الثالث (ر٣) يناظر المئيني ٧٥.

فالإرباعي الأول هو نقطة على التوزيع تقع دونها ٢٥٪ من الحالات، والإرباعي الثالث هو نقطة أخرى على التوزيع تقع دونها ٧٥٪ من الحالات.

أما المدى الربيعي Interquartile Range فهو عبارة عن مدى الدرجات التي تمثل ٥٠٪ الوسطى من الحالات في التوزيع، أي هو الفرق بين الإرباعي الثالث والإرباعي الأول.

وتعد الإعشاريات والإرباعيات حالات خاصة من المئينيات، وتستخدم الإرباعيات أيضًا في تحديد مركز الفرد بالنسبة لأقرانه.

فإذا قلنا أن درجة الفرد تناظر الإرباعي الأول فإننا نعني بذلك أن درجته تناظر نقطة معينة على ميزان الدرجات تقع دونها ٢٥٪ من الدرجات، كما تستخدم في معرفة الشكل العام لتوزيع الدرجات،وذلك بإيجاد الفرق بين القيمتين المناظرتين للإرباعي

الثالث والإرباعى الثانى، وكذلك الفرق بين الإرباعى الثانى والإرباعى الأول،أى (ر٣ - ر٢) ، (ر٢ - ر١).

فإذا كانت القيمتان متساويتان تقريبًا فإن التوزيع يكون متماثلاً ، أما إذا كانت إحداهما أكبر من الأخرى فإن التوزيع يكون ملتويًا.

ويمكن الإفادة من المدى الربيعى فى تعرف مدى تشتت توزيع الدرجات فى المنطقة الوسطى المحصورة بين ر١ ، ر٣ وبخاصة إذا كان التوزيع متماثلاً.

طرق حساب المئينيات :

نفترض أن لدينا التوزيع التكرارى التالى لدرجات ٤٠ فردًا فى أحد الاختبارات النفسية (جدول ٣ ـ ٦):

فئات الدرجات	التكرار	التكرار المتجمع	التكرار المتجمع المئوى	الحدود الحقيقية للفئات
٦٥ - ٦٩	٣	٣	٧,٥	٦٤,٥ - ٦٩,٥
٧٠ - ٧٤	٥	٨	٢٠,٠	٦٩,٥ - ٧٤,٥
٧٥ - ٧٩	٨	١٦	٤٠,٠	٧٤,٥ - ٧٩,٥
٨٠ - ٨٤	١١	٢٧	٦٧,٥	٧٩,٥ - ٨٤,٥
٨٥ - ٨٩	٧	٣٤	٨٥,٠	٨٤,٥ - ٨٩,٥
٩٠ - ٩٤	٤	٣٨	٩٥,٠	٨٩,٥ - ٩٤,٥
٩٥ - ٩٩	٢	٤٠	١٠٠,٠	٩٤,٥ - ٩٩,٥
	ن = ٤٠			

جدول (٣ ـ ٦) توزيع تكرارى لدرجات اختبار، وكذلك التكرار المتجمع للدرجات

ولحساب قيمة المئيني ٧٠ مثلا من جدول (٣ ـ ٦) دون اللجوء إلى استخدام صيغ رياضية معقدة، يمكن اتباع الخطوات التالية :

(١) نوجد عدد الأفراد الذين تقل درجاتهم عن المئيني ٥٠، وذلك بأن نضرب رتبة المئيني × العدد الكلي للأفراد أي ٠,٥٠ × ٤٠ = ٢٠.

(٢) نكوّن العمود الثالث الذي يشتمل على التكرار المتجمع وذلك بالبدء بالتكرار المقابل للفئة الأولى ونقوم بعملية جمع متتال للتكرارات التي في الفئات التي تليها.

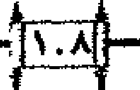
فمثلاً التكرار المتجمع الذي يناظر الحد الأعلى الحقيقي للفئة ٦٩,٥ - ٧٤,٥ نحصل عليه بجمع تكرار هذه الفئة على تكرار الفئة السابقة عليها أي ٣ + ٥ = ٨.

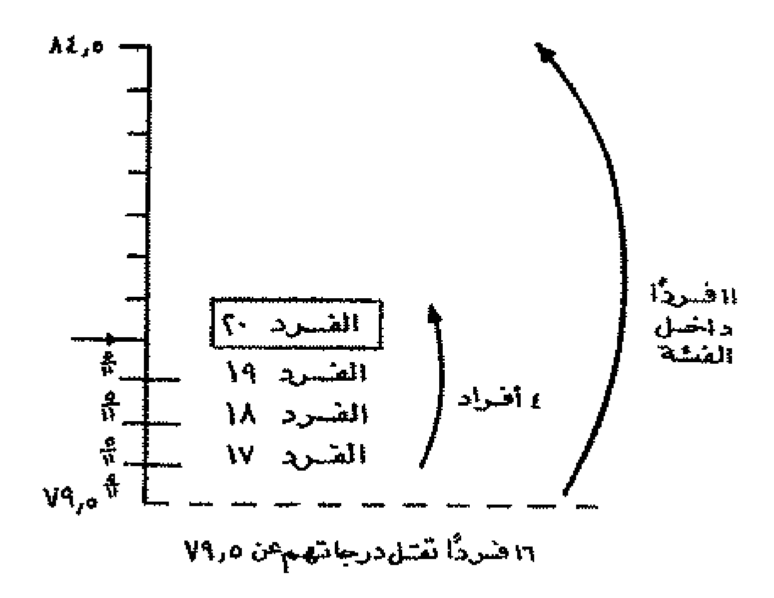
والتكرار المتجمع الذي يناظر الحد الأعلى الحقيقي للفئة ٧٤,٥ - ٧٩,٥ نحصل عليه بجمع تكرار هذه الفئة على تكرار الفئتين السابقتين أي ٣ + ٥ + ٨ = ١٦ وهكذا في بقية الفئات.

وينبغي التأكد من أن قيمة التكرار المتجمع للفئة الأخيرة ٩٤,٥ - ٩٩,٥ تساوي العدد الكلي للتكرارات. كما ينبغي ملاحظة أننا استخدمنا الحدود الحقيقية للفئات وذلك بإضافة ٠,٥ إلى الحدود الدنيا للفئات المدونة في العمود الأول، وطرح ٠,٥ من الحدود العليا لهذه الفئات لكي يصبح توزيع الدرجات متصلاً.

(٣) نبحث في عمود التكرار المتجمع (العمود الثالث) عن الحدود الحقيقية للفئة التي يقع فيها التكرار المتجمع ٢٠ المذكور في الخطوة الأولى، فنجد أن هذه الفئة هي : ٧٩,٥ - ٨٤,٥ ، ونلاحظ أن تكرار هذه الفئة المدون في العمود الثاني ١١. أي أن هناك ١١ فرداً ينتمون إلى هذه الفئة، فإذا افترضنا أن سعة الفئة التي تساوي ٥ درجات مقسمة إلى فترات متساوية على هؤلاء الأفراد، فإن كلا منهم سيشغل مسافة $\frac{٥}{١١}$ من هذه الفئة.

(٤) نظراً لأننا نبحث عن التكرار المتجمع ٢٠، فإننا عند التكرار المتجمع ١٦ نحتاج إلى ٤ مسافات داخل الفئة ٨٤,٥ - ٨٩,٥ للوصول إلى هذا التكرار المتجمع المطلوب، أي ٢٠ مطروحًا منها ١٦ ، ويمكن توضيح ذلك كما يلي :





(٥) نحسب الدرجة المناظرة للمسافات الأربع فنجدها $\frac{٤}{١١}$ × ٥ = $\frac{٢٠}{١١}$ = ١,٨

(٦) لإيجاد الدرجة المناظرة للمئينى ٥٠ نجمع الحد الأدنى الحقيقى لهذه الفئة أى ٧٩,٥ على الدرجة المناظرة للمسافات الأربع، أى ٧٩,٥ + ١,٨ = ٨١,٣ .

أى أن الدرجة المناظرة للمئينى ٥٠ = ٧٩,٥ + $\frac{٢٠ - ١٦}{١١}$ × ٥

= ٧٩,٥ + $\frac{٤}{١١}$ × ٥

= ٧٩,٥ + ١,٨ = ٨١,٣ تقريبًا

ومن ذلك يمكن أن نستنتج الصيغة العامة لإيجاد الدرجات المناظرة لرتب مئينية معينة حيث تكون كالتالى :

الدرجة المناظرة لرتبة مئينية معينة =

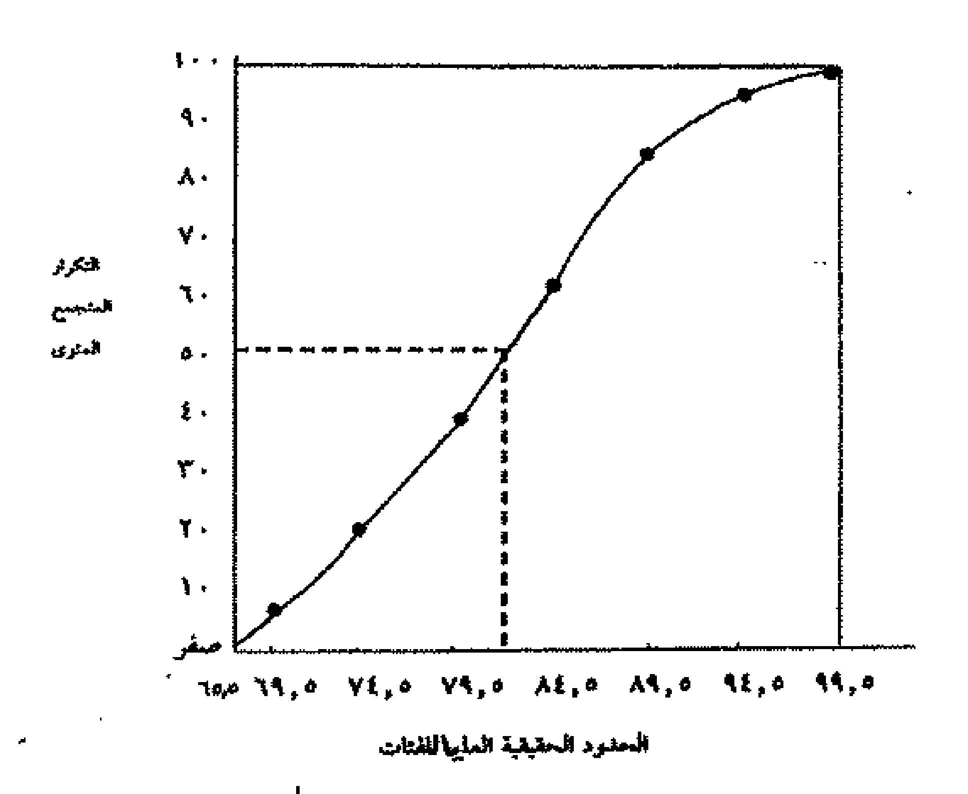
$$س_م + \frac{\text{التكرار المتجمع للدرجة المطلوبة} - \text{التكرار المتجمع تـم}}{\text{تكرار الفئة التى تحتوى التكرار المتجمع للدرجة}} \times \text{طول الفئة}$$
.... (٣ ـ ٩)

حيث ($س_م$) ترمز إلى الحد الأدنى الحقيقى للفئة التى تحتوى على التكرار المتجمع المناظر للرتبة المئينية المطلوبة.

، ($ت_م$) ترمز إلى التكرار المتجمع للفئة التى تحتوى على التكرار المتجمع المناظر للرتبة المئينية المطلوبة.

إيجاد الدرجات المقابلة للمئينيات بيانياً :

يمكن إيجاد الدرجات المقابلة للمئينيات باستخدام المنحنى التكرارى المتجمع المئوى حيث يتطلب ذلك تحويل التكرارات المتجمعة المقابلة للحدود العليا لفئات توزيع الدرجات إلى تكرارات متجمعة مئوية،وذلك بقسمة كل من هذه التكرارات المتجمعة على التكرار الكلى ونضرب الناتج × ١٠٠ كما هو مبين بالعمود الرابع فى جدول (٣ ـ ٦) ، ثم نمثل هذه الحدود العليا على المحور الأفقى والتكرار المتجمع المئوى على المحور الرأسى، ونرسم المنحنى التكرارى المتجمع المئوى كما هو موضح بشكل (٣ ـ ١٤) التالى:



شكل (٣ ـ ١٤) يوضح المنحنى التكرارى المتجمع المئوى لدرجات الاختبار المذكورة فى جدول (٣ ـ ٦)

فلإيجاد الدرجة المقابلة للمئيني ٥٠ مثلا، نحدد النقطة ٥٠ على المحور الرأسي، ونرسم خطًا أفقيًا من هذه النقطة حتى يلاقي المنحنى التكراري المتجمع المئوي في نقطة أخرى نسقط منها عمودًا على المحور الأفقي، فتكون نقطة الالتقاء مناظرة للدرجة المطلوبة وهي ٨١,٣ كما يتضح من شكل (٣ ـ ١٤)، ويلاحظ أنها مساوية للدرجة التي حصلنا عليها بالطريقة الأخرى.

ونظرًا لأن المئيني ٥٠ يناظر الوسيط، فإنه يمكن حساب الدرجات المقابلة لكل من الوسيط والإرباعيات بنفس هاتين الطريقتين، حيث إن الإرباعي الأول يناظر المئيني ٢٥، والإرباعي الثالث يناظر المئيني ٧٥ كما سبق أن أشرنا.

الرتب المئينية:

تستخدم الرتب المئينية Percentile Ranks بكثرة في تفسير درجات الاختبارات مرجعية الجماعة أو المعيار Norm - Referenced Tests وذلك لسهولة توضيح المقصود منها للطلاب والمفحوصين وأولياء الأمور حيث إنها لا تتطلب منهم إلا معرفة مفهوم النسبة المئوية.

ففي هذا النوع من الاختبارات نود معرفة النسبة المئوية لعدد الأفراد الذين يحصلون على درجات تقل قيمتها عن قيمة درجة معينة في أحد الاختبارات. فإذا كانت الرتبة المئينية المقابلة للدرجة الخام ٥٧ هي ٦٠، فإن هذا يعني أن ٦٠٪ من أفراد مجموعة معينة تقل درجاتهم عن الدرجة ٥٧، بينما تزيد درجة ٤٠٪ منهم عن هذه الدرجة، فالرتبة المئينية إذن لا يمكن تفسيرها إلا في إطار جماعة مرجعية معينة Reference Group يطبق عليهم الاختبار أو المقياس التربوي أو النفسي. فإذا حصل طالب على درجة خام تقابل رتبة مئينية ٨٥ في اختبار تحصيلي، فربما يبدو لأول وهلة أن أداء الطالب مرتفعًا لأنه يتفوق على ٨٥٪ من أقرانه، ولكن إذا كانت الجماعة المرجعية التي نقارنه بها من ضعاف التحصيل بعامة، فإن أداءه في هذه الحالة يعد منخفضًا. وعلى العكس من ذلك، إذا كانت جماعته المرجعية من الموهوبين، وحصل على درجة خام تقابل رتبة مئينية ٣٥ مثلا، فإن أداءه ربما يبدو جيدًا، وهكذا. لذلك ينبغي الحيطة عند المقارنة بين الرتب المئينية لجماعات مرجعية مختلفة لكي لا يساء تفسير هذه الرتب.

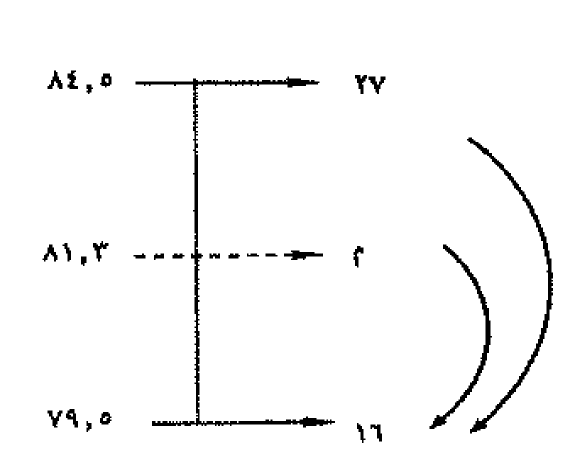
طرق حساب الرتب المئينية :

نظراً لأن الرتب المئينية تشير إلى نسبة من الأفراد الذين يحصلون على درجات أقل من درجة معينة فى اختبار ما، فإن هذه الدرجة تمثل نقطة تسمى المئينى، أى أن الرتبة المئينية عكس المئينى، ولذلك فإن طرق حسابها لا تختلف كثيراً عن طرق حساب المئينيات التى ذكرناها.

ولتوضيح هذه الطرق نعود مرة أخرى إلى توزيع الدرجات الموضح فى جدول (٣ ـ ٦).

ونفترض أننا نود إيجاد الرتبة المئينية المقابلة للدرجة الخام ٨١,٣ ، فما علينا إلا أن نحدد الفئة التى تقع فيها هذه الدرجة حيث نجدها الفئة ٧٩,٥ - ٨٤,٥ ، ثم نبحث فى العمود الرابع عن التكرار المتجمع المئوى لكل من الدرجتين ٧٩,٥ ، ٨٤,٥ فنجد أنه ٤٠ ، ٦٧,٥ على الترتيب.

وهذا يعنى أن الرتبة المئينية المقابلة للدرجة ٨١,٣ تقع بين ٤٠ ، ٦٧,٥ كما هو موضح فيما يلى :



ونستطيع الآن تحديد الرتبة المئينية (م) المقابلة للدرجة ٨١,٣ باستخدام النسبة والتناسب كالتالى:

$$\frac{\text{م} - \text{٤٠}}{\text{٦٧,٥} - \text{٤٠}} = \frac{\text{٨١,٣} - \text{٧٩,٥}}{\text{٨٤,٥} - \text{٧٩,٥}}$$

أى أن :

$$\frac{\text{م} - \text{٤٠}}{\text{٢٧,٥}} = \frac{\text{١,٨}}{\text{٥}}$$



$$5(م - 40) = 1,8 \times 27,5$$

$$5م - 200 = 49,5$$

$$5م = 249,5$$

أى أن : م = $\frac{249,5}{5}$ = 49,9 أى تساوى 50 تقريبًا.

وهى نفس الرتبة المذكورة فى المثال السابق.

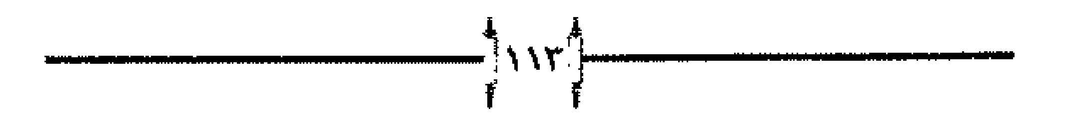
ويمكن الحصول على هذه الرتبة بيانيًا باستخدام شكل (3 ـ 14) بطريقة عكسية حيث نبدأ بالمحور الأفقى (المحور الذى يمثل الحدود الحقيقية العليا لفئات الدرجات) ونحدد النقطة 81,3 على هذا المحور، ونرسم منها خطا رأسيًا يلاقى المنحنى التكرارى المتجمع المئوى، ثم نرسم عمودًا على المحور الرأسى (المحور الذى يمثل التكرار المتجمع المئوى أو الرتب المئينية) من نقطة التلاقى، فتكون نقطة تقاطع العمود الأفقى مع المحور الرأسى هى الرتبة المئينية المطلوبة، ونلاحظ أنها تساوى 50.

بعض أوجه قصور المئينيات والرتب المئينية:

على الرغم من أهمية المئينيات والرتب المئينية وسهولة استخدامها فى تفسير درجات الاختبارات والمقاييس وفى تحديد المركز النسبى للفرد فى إطار جماعة مرجعية معينة طبق عليها الاختبار أو المقياس، إلا أنه يشوبها بعض أوجه القصور فى هذا الشأن ونوجزها فيما يلى:

(1) الرتب المئينية تتوزع توزيعًا مستطيلاً فى حين أن توزيعات درجات الاختبارات وبخاصة الاختبارات مرجعية الجماعة تتخذ عادة شكل المنحنى الاعتدالى. ولذلك فإن القيم العددية المتساوية للفروق بين الرتب المئينية لا تعنى فروقًا متساوية في الدرجات الخام، فالفرق بين الرتبة المئينية 32 والرتبة المئينية 38 يساوى 4 نقاط، بينما الفرق بين الدرجتين الخام المقابلتين لهما ربما تكون 10 درجات، وربما تكون أقل بكثير. فمجرد النظر إلى الرتب المئينية لا يكفى لمعرفة الفروق بين الدرجات الخام. كذلك فإن الفروق الضئيلة بين الدرجات الخام بالقرب من مركز التوزيع تناظر رتبا مئينية كبيرة، بينما الفروق الكبيرة بين الدرجات الخام عند طرفى التوزيع تناظر فروقا صغيرة فى هذه الرتب.

(2) تتوزع الرتب المئينية على ميزان رتبى، ولذلك لا نستطيع إجراء العمليات الحسابية الأساسية عليها، غير أن هذا القصور لا يؤثر فى تفسير درجات الاختبارات.



وقد سبق أن أوضحنا أن الدرجات المعيارية يمكن ضمّها بحيث تشكّل درجة معيارية مركبة حيث إنها تتوزع على ميزان فترى.

ولكن إذا كان توزيع الدرجات الخام اعتداليًا فإنه يمكن الإفادة من خصائص المنحنى الاعتدالى التى سبق توضيحها فى إيجاد المئينيات المناظرة للدرجات المعيارية، وبذلك تجعل تفسير درجات الاختبار أكثر سهولة؛ لأنه كما ذكرنا يصعب على الطالب أو المفحوص العادى تفسير الدرجات المعيارية.

استخدام المنحنى الاعتدالى فى إيجاد المئينيات :

لتوضيح كيفية استخدام المنحنى الاعتدالى فى إيجاد المئينيات دعنا ننظر إلى المنحنى الاعتدالى الموضح بشكل (٣ - ١٣) ، ونفترض أن توزيع الدرجات الخام فى أحد الاختبارات كان توزيعًا اعتداليًا متوسطه ١٠٠ وانحرافه المعيارى ١٦، وأردنا تحديد الرتبة المئينية المقابلة للدرجة الخام ١٣٢ باستخدام خصائص المنحنى الاعتدالى.

فالخطوة الأولى : نحول الدرجة الخام إلى درجة معيارية باستخدام الصيغة (٣-٦) فنجد أنها تساوى $\frac{١٣٢ - ١٠٠}{١٦} = + ٢$.

والخطوة الثانية : نحدد المساحة تحت المنحنى الاعتدالى إلى يسار هذه الدرجة، فنجد أنها تساوى ٩٨ تقريبًا، وهذا يعنى أن الدرجة ١٣٢ تقابل الرتبة المئينية ٩٨، أى أن هذه الدرجة تفوق ٩٨٪ من درجات الجماعة المرجعية.

وعلى العكس من ذلك فإننا نستطيع إيجاد الدرجة الخام المقابلة لرتبة مئينية معينة بأن نتبع الخطوة الثانية أولاً ثم الخطوة الأولى. وينبغى التنويه إلى أنه إذا كانت الدرجة المعيارية كسرية، فإنه توجد جداول إحصائية تحدد المساحات تحت المنحنى الاعتدالى فى هذه الحالة، ويمكن الرجوع إلى هذه الجداول فى أحد المراجع الأولية فى الإحصاء.

كما ينبغى التنويه إلى أنه إذا كان توزيع الدرجات الخام فى الاختبار يبتعد عن التوزيع الاعتدالى، فإنه يمكن تحويل هذا التوزيع إلى توزيع اعتدالى باستخدام طرق معينة لا مجال لذكرها هنا، وقد قام المؤلف بتوضيح هاتين الحالتين فى مرجع آخر.

فإذا أجرى هذا التحويل على توزيع الدرجات الخام يصبح شكل توزيع الدرجات اعتداليًا بغض النظر عن شكل التوزيع الأصلى للدرجات الخام. فإذا حوّلت هذه الدرجات الجديدة إلى درجات معيارية، فإن الدرجات المعيارية الناتجة تسمى الدرجات

المعيارية الاعتدالية Normalized Z ، وكذلك إذا حُوِّلت الدرجات المعيارية الاعتدالية إلى درجات تائية ، فإن الدرجات التائية الناتجة تسمى الدرجات التائية الاعتدالية Normalized T.

فإذا علمنا قيمة كل من الدرجة المعيارية الاعتدالية والدرجة التائية الاعتدالية، فإن العلاقة بين هاتين القيمتين وقيمتى المئينيات المقابلة لهما يحددها المنحنى الاعتدالى المعيارى الموضح بشكل (٣ - ١٣) بغض النظر عن شكل توزيع الدرجات الخام.

لذلك إذا علمنا أن الدرجة المعيارية الاعتدالية لفرد في اختبار ما (-١) مثلاً، فإن هذا يعنى أن هذه الدرجة تقابل المئينى ١٦ تقريبًا.

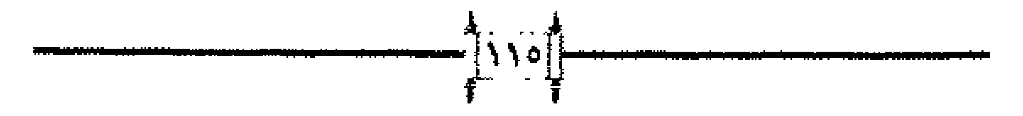
وبالطبع لا يجوز مثل هذا التفسير إذا كانت الدرجة المعيارية الخطية المعتادة (-١) ما لم يكن توزيع الدرجات الخام اعتداليًا.

مقاييس العلاقة بين مجموعتين من الدرجات :

لاحظنا مما سبق أن مقاييس النزعة المركزية أو الموضع، ومقاييس التشتت، ومقاييس المركز النسبى،مثل الدرجات المعيارية وغيرها من الدرجات المعيارية المعدلة، والمئينيات والرتب المئينية،تستخدم في وصف توزيعات درجات الاختبارات والمقاييس، وتحديد المركز النسبي للفرد في إطار جماعتـه المرجعية. وجميع هذه المقاييس تصف متغيـرا واحداً. ولكننا فى القياس التربـوى والنفسى نهتم فى كثيـر من الأحيان بوصف العلاقة بين متـغيرين أو سمتين من سمـات الأفراد. فإذا طبق اختباران يقيسـان سمتين مختلفتين فإننا نحصل على مجموعتين من الدرجات.وهنا ربما نهتم بمعرفة العلاقة بين المراكز النسبية للأفراد في إحدى السمتين ومراكزهم النسبية فى السمة الأخرى.

فمثلاً ربمـا نود معرفة العلاقة بين اختبارين أحدهما يقيس الاستعداد الدراسى والآخر يقيس التحصيل، أو بين مقياسين من مقاييس الذكاء أو مقاييس الشخصية، وهكذا، ففي هذه الحالات ربما نجد أن الأفراد الذين حصلوا على درجات مرتفعة فى أحد الاختبارين أو المقياسين يميلون إلـى الحصول على درجات مرتفعة فى الاختبار أو المقياس الأخر، فهنا تكون العلاقة بين مجموعتى الدرجات موجبة. وعلى العكس من ذلك، إذا كانت درجاتهم فى أحد الاختبارين منخفـضة مثلا بينما درجاتهم فى الاختبار الأخر تميل إلى الارتفاع فإن العلاقة بين مجموعتى الدرجات تكون سالبة.

ومقـاييس العـلاقة تحـدد قيمـة عددية تمـثل درجة الارتـباط بين مـجموعـتى الدرجات، أى إلى أى مدى يحـافظ الأفراد على مراكزهم النسبية فى كل من السمتين موضع القيـاس.



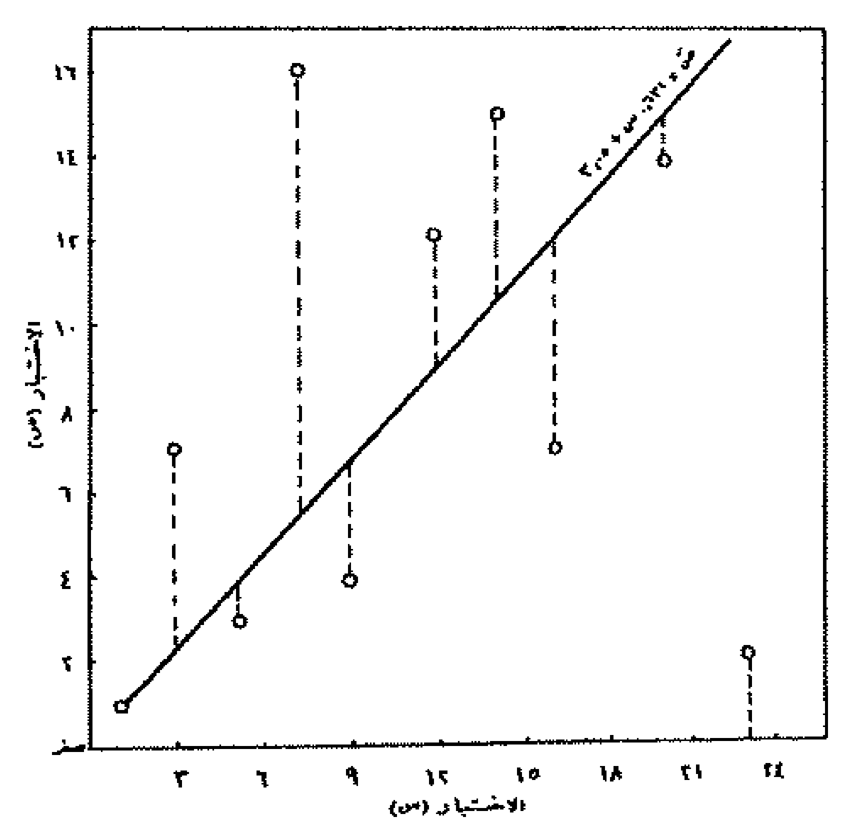
ويمكن أن تتضح مثل هذه العلاقات إذا قمنا بتمثيل قيم المتغيرين أى مجموعتى الدرجات تمثيلاً بيانيًا فى شكل واحد يسمى الشكل الانتشارى Scatter Diagram للمتغيرين.

فإذا افترضنا أن لدينا مجموعتى الدرجات التالية لعينة من الطلاب فى اختبارين أحدهما يقيس التحصيل فى الرياضيات والآخر يقيس التحصيل فى الفيزياء (جدول٣ـ٧):

اختبار الرياضيات (س)	٢٣	٢٠	١٦	١٤	١٢	٩	٧	٥	٣	١
اختبار الفيزياء (ص)	٢	١٤	٧	١٥	١٣	٤	١٦	٣	٧	١

جدول (٣ ـ ٧) درجات عينة من الطلاب فى اختبارين تحصيليين

فإنه يمكن تمثيل قيم (س) على المحور الأفقى ، وقيم (ص) على المحور الرأسى، ونحدد لكل زوج من القيم المتناظرة لكل من (س) ، (ص) نقطة على مستوى المحورين كما بشكل (٣ ـ ١٥) التالى :



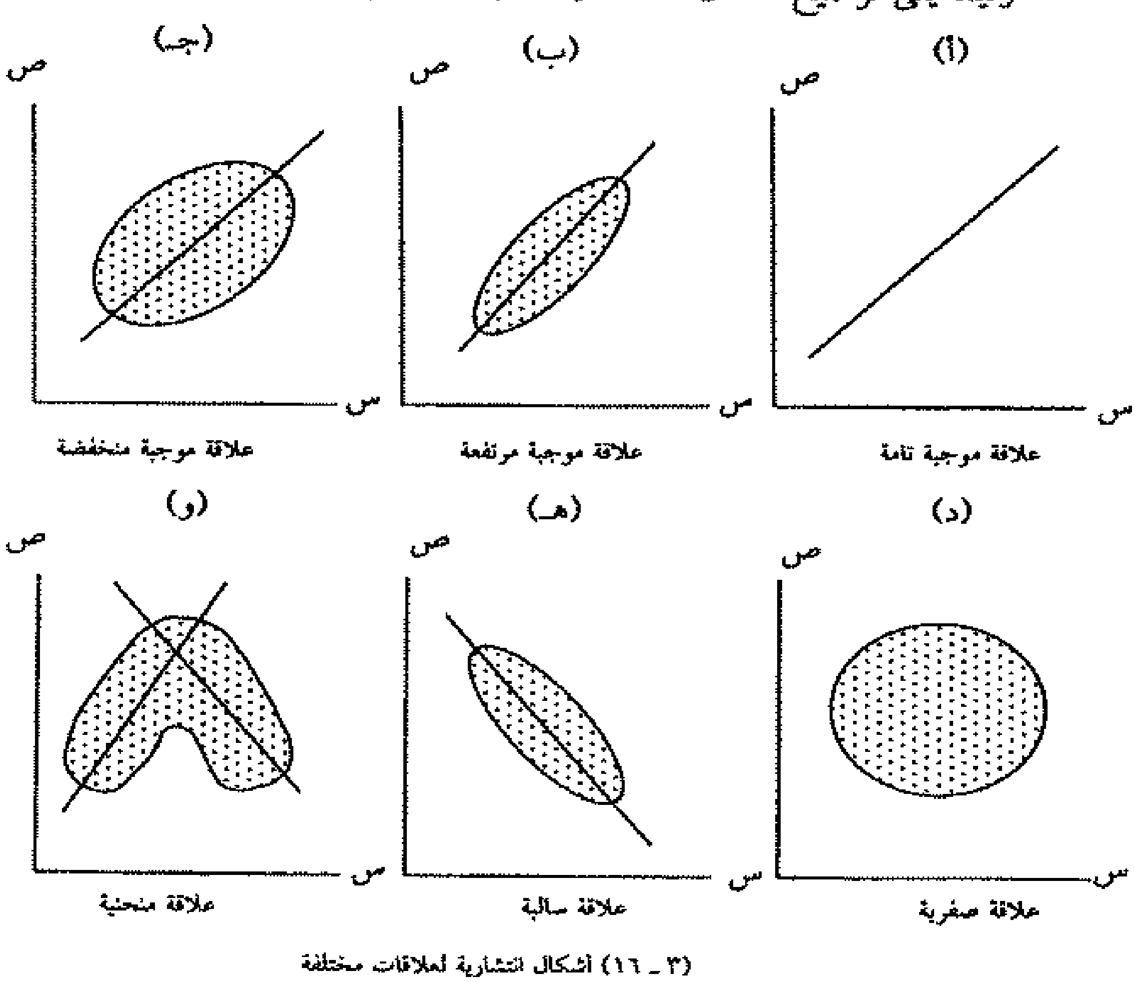
شكل (٣ ـ ١٥) شكل انتشارى لأزواج الدرجات المذكورة فى جدول (٣ ـ ٧)

ويتضح من شكل (٣ ـ ١٥) أنه على الرغم من انتشار النقط التى تمثل أزواج درجات الاختبارين، إلا أنها تقع جميعـها داخل شكل بيضاوى، وهناك نزعة عامة لأن تقع هذه النقـط على خط مستقـيم يمر بمعظـمها. لذلك فإن العـلاقة بين مجمـوعتى الدرجات يقال أنها علاقة خطية أو مستقيمة Linear Relationship . ونظراً لأن الشكل البيضاوى يتجه إلى أعلى جهة اليمين، فإن العلاقة موجبة.

لذلك فإن هذا الشـكل الانتشارى يفيد فـى إلقاء الضوء على شكل العـلاقة بين مجموعتى الدرجات لمعرفة ما إذا كانت العلاقة مستقيمة أم منحنية، وموجبة أم سالبة.

فالعلاقة المـنحنية لا تميل فيها الـنقط إلى التراكم حول خط مستـقيم، بل تبدو هناك نزعة للانحناء، أى يكون هناك نزعة للاتجـاه إلى أعلى فى جزء معين ثم ينعكس الاتجاه إلى أسفل في جزء آخـر من الشكل. أما العلاقة السالبة فتـبدو من اتجاه الشكل البيضاوى إلى أسفل جهة اليمين .

وفيما يلى توضيح أشكال هذه العلاقات (٣ ـ ١٦) :



(٣ ـ ١٦) أشكال انتشارية لعلاقات مختلفة

تحديد مقدار العلاقة بين مجموعتين من الدرجات :

لاحظنا أن الأشكال الانتشارية توضح شكل العلاقة ونوعها ولكن لا تحدد مقدار هذه العلاقة تحديدًا كميًّا. فلإيجاد مقدار واتجاه العلاقة بين مجموعتين من الدرجات في حالة ما إذا كانت العلاقة خطية أو مستقيمة كما في الأشكال أ ، ب ، جـ ، د يمكن استخدام مقاييس إحصائية مناسبة. ولعل أهم هذه المقاييس معامل ارتباط بيرسون Pearson Product Moment Correlation . ولهذا المعامل صيغ رياضية متعددة بعضها يعتمد على الدرجات المعيارية أو الانحرافات عن المتوسط، والبعض الآخر يعتمد على الدرجات الخام مباشرة.

ونظرًا لأن الصيغة الأخيرة يمكن تطبيقها بسهولة باستخدام آلة حاسبة صغيرة أو الحاسوب، فإننا سوف نقتصر على هذه الصيغة لأنها شائعة الاستخدام. ونقدم برنامجا بسيطًا متعلقا بها يمكن تنفيذه بسهولة باستخدام الحاسوب.

معامل ارتباط بيرسون :

لإيجاد العلاقة بين مجموعتين من الدرجات باستخدام معامل ارتباط بيرسون ينبغي التحقق من استقامة العلاقة من الشكل الانتشاري لمجموعة الدرجات، وكذلك التحقق من أن السمة مقاسة على ميزان فترى على الأقل.

والصيغة العامة لمعامل ارتباط بيرسون بين متغيرين (س) ، (ص) كالتالي:

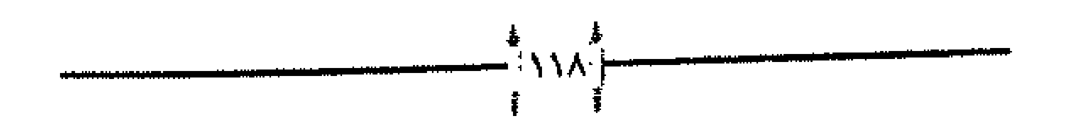
$$\text{ر} = \frac{\text{ن مجـ س ص} - \text{مجـ س مجـ ص}}{\sqrt{[\text{ن مجـ س}^{2} - (\text{مجـ س})^{2}][\text{ن مجـ ص}^{2} - (\text{مجـ ص})^{2}]}}$$

(٣-١٠).

ويتطلب تطبيق هذه الصيغة إيجاد قيم كل من :

(مجـ س ص) ،(مجـ س × مجـ ص) ، (مجـ س ٢)، (مجـ س) ٢ ، (مجـ ص ٢) ، (مجـ ص) ٢

ويوضح جدول (٣ ـ ٨) كيفية تطبيق هذه الصيغة بمثال لمجموعتي الدرجات (س) ، (ص) المشار إليهما في جدول (٣ ـ ٧) والشكل الانتشاري (٣ ـ ١٥):



س	ص	س²	ص²	س ص
٢٣	٢٠	٥٢٩	٤٠٠	٤٦٠
٢٠	١٤	٤٠٠	١٩٦	٢٨٠
١٦	٧	٢٥٦	٤٩	١١٢
١٤	١٥	١٩٦	٢٢٥	٢١٠
١٢	١٣	١٤٤	١٦٩	١٥٦
٩	٤	٨١	١٦	٣٦
٧	١٦	٤٩	٢٥٦	١١٢
٥	٣	٢٥	٩	١٥
٣	٧	٩	٤٩	٢١
١	١	١	١	١

جدول (٣ ـ ٨) يوضح خطوات إيجاد قيمة معامل ارتباط بيرسون

مجـ س = ١١٠ ، مجـ ص = ١٠٠ ، مجـ س^2 = ١٦٩٠

مجـ ص^2 = ١٣٧٠ ، مجـ س ص = ١٤٠٣

$$\text{ر} = \frac{١٠ \times ١٤٠٣ - ١١٠ \times ١٠٠}{\sqrt{[١٠ \times ١٦٩٠ - (١١٠)^2][١٠ \times ١٣٧٠ - (١٠٠)^2]}}$$

$$= \frac{٣٠٣٠}{٤٨٠٠ \times ٣٧٠٠} = \frac{٣٠٣٠}{٤٢١٤,٣} = ٠,٧١٩$$

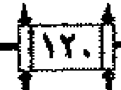
وينبغي ملاحظة أن معامل ارتباط بيرسون تتراوح قيمه بين (-١) ، (+١) ، فالقيمة (-١) تعني ارتباط تام سالب، والقيمة (+١) تعني ارتباط تام موجب ، والقيمة (صفر) تعني عدم وجود ارتباط بين المتغيرين. وعادة تكون قيم معامل ارتباط

بيرسون بين مجموعتين من درجات الاختبارات والمقاييس التربوية والنفسية التى تستخدم فى دراسة الفروق الفردية أقل من الواحد الصحيح.

برنامج الحاسوب الشخصي أو حاسوب الجيب لإيجاد معامل ارتباط بيرسون :

يمكن استخدام برنامج الحاسوب التالى فى إجراء العمليات الحسابية التى تتطلبها الصيغة (٣ ـ ١٠) . ويتميز هذا البرنامج بالبساطة وسرعة الإجراء، علما بأننا سوف نرمز إلى المتغير (س) بالرمز (x) ، والمتغير (ص) بالرمز (y) ، ومجموع قيم س بالرمز S_X ، ومجموع قيم ص بالرمز S_y، ومجموع مربعات قيم س بالرمز S_{XX}، ومجموع مربعات قيم ص بالرمز S_{yy} ، ومجموع حواصل ضرب قيم س فى قيم ص S_{xy}.

```
5     N = 0 :  SX  = 0 : Sy = 0 : XY = 0.
10    Read X , Y
20    if X <  - 999 THEN  110
30    N = N + 1
40    Sx =  Sx + X
50    Sy = Sy + Y
60    XX = XX + X * X
70    YY = YY + Y * Y
80    XY = XY + X * Y
100   GO TO 10
110   Rem Calc B And A
120   B = ( N * XY - Sx * Sy ) / N * XX - Sx * Sx)
130   A = ( Sy - B *  Sx ) / N
140   PRINT A , B
150   R = ( N * XY - Sx* Sy ) / SQR ( N * XX - Sx * Sx ) * ( N
      * YY - Sy * Sy )
160 PRINT R
```



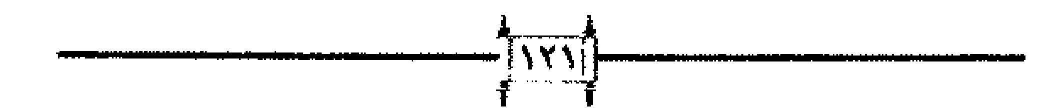
كما أن هذا البرنامج يوجد قيمة كل من الثابتين A, B اللذين يمكن استخدامهما في إيجاد معادلة الانحدار التي سنوضحها بعد قليل.

تفسير قيم معامل ارتباط بيرسون :

إن قيم معامل ارتباط بيرسون لا يجب تفسيرها على أنها نسب مئوية، فلا ينبغي أن نكتب القيمة ٧١٩, ٠ لمعامل الارتباط ٧٢ ٪ ، وإنما يمكن تقريبها إلى رقمين عشريين على الأقل، أي يمكن كتابتها ٧٢, ٠، ولا يجوز كتابتها ٧, ٠ تحريا للدقة. كما لا ينبغي إجراء عمليات جمع أو طرح على قيم معاملات الارتباط نظراً لأنها معاملات من المستوى الرتبي،فالفرق بين معاملي ارتباط ٣٠, ٠ ، ٤٠, ٠ لا يمثل نفس الفرق بين معاملي الارتباط ٨٠, ٠ ، ٩٠, ٠ مثلا.

ولذلك لا ينبغي تفسير قيم معاملات الارتباط على أنها تشير إلى علاقات علية أو سببية ، وإنما تعني فقط وجود اتجاه ومقدار معين لهذه العلاقات، فالعلاقة بين الذكاء والتحصيل لا تعني بالضرورة أن الذكاء يسبب التحصيل ، حيث إن هناك عوامل أخرى كثيرة تؤثر في التحصيل، فمعاملات الارتباط لا تتضمن بالضرورة علاقات علية. كما يجب عند تفسير قيم معاملات الارتباط مراعاة عدد أفراد العينة المستخدمة في حساب هذه القيم حيث إن هذه القيم، تتأثر تأثراً ملحوظاً باختلاف عدد أفراد العينة. فمعامل الارتباط يعتمد على النزعة المركزية والتشتت لكل من المتغيرين، وكلما زاد التشتت زادت قيمة معامل الارتباط والعكس صحيح. ومما لا شك فيه أن عدد أفراد العينة يؤثر في تشتت الدرجات أي تباين هذه الدرجات، مما يؤثر بدوره في قيم معامل ارتباط بيرسون.

ولتفسير قيم معامل الارتباط ينبغي تربيعها للحصول على معامل آخر يسمى معامل التحديد Coefficient of Determination ، وهذا المعامل يحدد نسبة التباين المشترك بين المتغيرين، أي النسبة المئوية للتباين في درجات المتغير الأول التي تعزى إلى تباين درجات المتغير الثاني. فمعامل الارتباط الذي قيمته ٧١٩, ٠ والتي حصلنا عليها في المثال السابق لا نستطيع القول بأنها قيمة مرتفعة ؛لأن تربيع هذه القيمة يجعلها مساوية ٥٢, ٠ تقريباً، مما يدل على ٥٢٪ من تباين درجات اختبار الفيزياء تعزى إلى تباين درجات اختبار الرياضيات، أما بقية التباين وهو ٤٨٪ فيعزى إلى عوامل أخرى لا نستطيع تحديدها بدقة، إذ ربما تكون الذكاء أو الدافعية أو عوامل مدرسية معينة، ولذلك فهي قيمة متوسطة.



وعلى الرغم من أن قيم معامل الارتباط لا يجوز المقارنة بينهـا ، إلا أن القيم النسبية لمعامل التحديد المناظرة لـها يمكن المقـارنة بينها مقـارنة مباشرة، فمعامل الارتباط ٠,٧١٩ لا نستطيع اعتباره ضعف معامل الارتباط ٠,٣٦٠ ، حيث إن معامل التحديد المنـاظر لكل منهما ٠,٥٢ ، ٠,١٣ على الترتيب، ونـظرا لان القيمة ٠,٥٢ تساوى أربعة أضعاف القيمة ٠,١٣ فإن قوة معامل الارتباط ٠,٧١٩ تساوى ضعفى قوة معامل الارتباط ٠,٣٦٠ ، وليس كما يبدو أنها الضعف فقط.

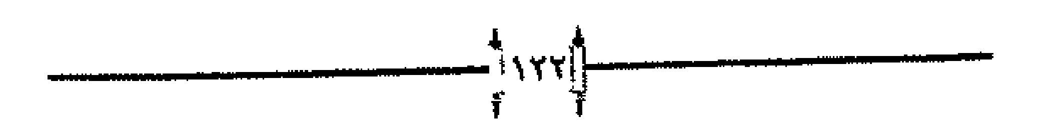
والخلاصة أنـه ينبغى مراعاة شـروط استخدام معـامل ارتباط بيرسـون من حيث ميزان القياس الفترى لدرجات الاختبارات، واستـقامة العلاقة بين مجموعتى الدرجات. فإذا كان ميزان القيـاس اسميًا أو رتبيًا أو أن العلاقة منحنية فـإنه لا يجوز استخدام هذا المعامل، حيث إن هنـاك مقاييسًا إحصائية أخرى تناسـب هذه الحالات يمكن الرجوع إليها فى المراجع الإحصائية.

اعتبارات أخرى عند تفسير قيم معامل الارتباط:

إن إساءة استخدام معامل الارتباط فى مجالات القياس التربوى والنفسى سواء فى إعداد الاختبارات أو فى تفسير درجاتها يؤدى إلـى نتائج تفتقر إلى الدقة، ولذلك ينبغى إلى جانب الملاحظات السابقة مراعاة بعض الاعتبارات الأخرى نوجزها فيما يلى :

(١) القيم المقبولة لمعامل الارتباط: فكثيراً ما يسيء كثير من المشتغلين بالقياس تفسير قيم معامل الارتباط كما سبق أن ذكرنا. كما أن البعض يتساءل عن القيم المقبـولة لمعامل الارتباط، فهل ٠,٦٠ أو ٠,٧٠ قـيمتان مقبـولتان أم لابد أن تكون قيمة معـامل الارتباط ٠,٨٥ أو ٠,٩٠ مثلاً؟ والحقيقـة أن اعتبار قيمة معـينة لمعامل الارتباط مرتفـعة يحتاج إلى أخذ كثيـر من العوامل بعين الاعتبـار، إذ ليس هناك نقطة محددة تفصل بيـن القيم المرتفعة والقيم المـنخفضة لهذا المعامل، وإنمـا يعتمد ذلك على طبيعة الاختبار أو المقياس المستخـدم، والمتغير أو السمة التى يقيسها، والعوامل المؤثرة فى هذه السمة.

ولذلك فإن القيمة المرتفعة أو المنخفضـة لمعامل الارتباط لا تعتمد اعتمادا كليًا على القيمة العددية لهذا المعامل، وإنما تعتمـد أيضًا على طبيعة الاختبار وفيم تستخدم نتائجه، فهذه القيمة العددية تشير إلى درجة العلاقة وليس إلى كيفية تفسير هذه العلاقة. وربما يُمكـن ملاحظة هذه المشكـلة بوضوح فى الاختبـارات التى تستخـدم فى مجال الانتقاء، حيث ترتبط درجات الاختبارات التـنبؤية بمحكات النجاح فـى دراسة أو مهنة



معينة، وتبلغ عادة قيمة معامل الارتباط ٠,٦٠ أو أكثر. ومع هذا فإن انخفاض قيمة هذا المعامل عن ٠,٦٠ في بعض هذه الاختبارات لا يقلل من شأن الاختبار حيث إنه يظل مفيداً في الانتقاء عما لو تم الانتقاء عشوائيًا.

(٢) عوامل الخطأ التي تؤثر في قيم معامل الارتباط : فقيم معامل الارتباط التي يدونها الباحثون للاختبارات التي يقومون بإعدادها، وكذلك المدونة في أدلة الاختبارات يتم تفسيرها على أنها توضح مقدار العلاقة بين السمات التي تقيسها هذه الاختبارات.

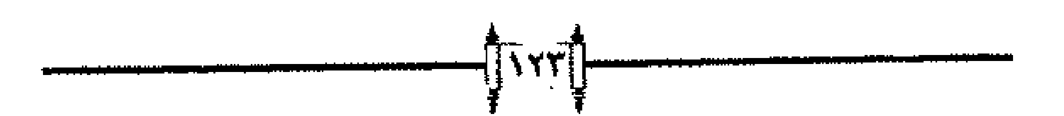
غير أن هناك عوامل خطأ كثيرة ربما تكون قد أسهمت في زيادة هذه القيم أو انخفاضها. ومن بين هذه العوامل المؤثرة : الأخطاء العشوائية، وضيق مدى درجات الاختبار، وشكل توزيع الدرجات، وعدم استقامة العلاقة بين مجموعتي الدرجات. فهذه العوامل تؤثر بلا شك تأثيراً متباينًا في قيم معامل الارتباط ، فالأخطاء العشوائية في أي من الاختبارين أو اختيار مجموعة متجانسة من الأفراد في متغيرات معينة، مثل العمر أو الذكاء أو التحصيل المرتفع يؤدي إلى انخفاض قيمة معامل الارتباط.

كما أن انحناء العلاقة أي بعدها عن الاستقامة تؤدي إلى قيم مضللة أو زائفة Spurious Correlation إذا استخدم معامل ارتباط بيرسون.

(٣) شكل توزيع كل من المتغيرين ،لكي تصل قيمة معامل ارتباط بيرسون إلى أقصى قيمتيها وهما (-١) ، (+١) يجب أن يكون توزيعا المتغيرين لهما نفس الشكل تقريبًا. فمثلاً إذا كان أحد المتغيرين متصلاً والآخر ثنائيًا (أي يتخذ قيمتين فقط مثل واحد صحيح أو صفر)، أو أن أحد التوزيعين ملتويًا التواء موجبًا والآخر ملتويًا التواءً سالبًا، فإن قيمة معامل الارتباط في كل من هاتين الحالتين تكون أقل من الواحد الصحيح.

استخدام معامل الإرتباط في التنبؤ :

لا يقتصر استخدام معامل الارتباط على وصف درجة العلاقة بين متغيرين فحسب بل يمكن استخدامه أيضًا في التنبؤ بقيمة أحد المتغيرين بمعلومية قيمة المتغير الآخر. فمثلاً إذا كان هناك ارتباط موجب بين اختبار يقيس الاستعداد الدراسي لدى الطلاب الذين يودون الالتحاق بالجامعة ومتوسط تقديراتهم في نهاية السنة الجامعية الأولى ، فإن معامل الارتباط يوضح درجة العلاقة بين درجات الاختبار ومتوسط التقديرات، وكذلك يمكن استخدامه في التنبؤ بمتوسط تقديرات الطلاب إذا علمنا درجاتهم في اختبار الاستعداد الدراسي. وعلى الرغم من أن كثيرًا من مقاييس الجوانب



المعرفية ترتبط فيما بينها بعلاقة خطية أى مستقيمة، إلا أنه توجد بعض المقاييس وبخاصة المتعلقة بالمتغيرات الفسيولوجية والدافعية ترتبط فيما بينها بعلاقة متحنية. ولذلك فإننا سوف نقتصر فى دراسة التنبؤ على العلاقات المستقيمة.

فإذا عدنا إلى شكل (٣ ـ ١٥) نجد أن معظم النقط تتراكم حول خط مستقيم يمثل النزعة العامة للدرجات. فإذا كان هذا الخط يجعل مجموع مربعات المسافات المبينة بالشكل الانتشارى الموازية للمحور الرأسى من النقط إلى الخط المستقيم أقل ما يمكن. فإن هذا الخط عندئذ يسمى خط الانحدار Regression Line ،وهو الخط الذى يستخدم فى التنبؤ بقيم المتغير (ص) من قيم المتغير (س) المعلومة بأقل قدر ممكن من الخطأ الذى يسمى خطأ التنبؤ Prediction Error . وهذا الخطأ عبارة عن الفرق بين الدرجة الملاحظة (ص) والدرجة المتنبأ بها صَ ، ونظرًا لأن مقدار هذا الخطأ يكون كبيرًا نسبيًّا فى القياس التربوى والنفسى، فإنه ينبغى العناية بانتقاء أدوات القياس التى تتنبأ بالظاهرة التى تهتم بها بأقل قدر ممكن من أخطاء التنبؤ.

وربما يتساءل البعض عما إذا كان من الضرورى أو من المطلوب أن نتنبأ بقيمة متغير (ص) من متغير معلوم (س) عندما تكون قيم كل من المتغيرين معلومة. وللإجابة عن ذلك ينبغى أن نعلم أن الاختبارات التى تستخدم فى التنبؤ تطبق فى بادئ الأمر على عينة عشوائية كافية وممثلة للمجتمع المستهدف وإيجاد الارتباط بين درجاتها وبين قيم المتغير المراد التنبؤ به، والتوصل إلى خط الانحدار الذى يستخدم مستقبلاً على عينات مماثلة مستمدة من المجتمع نفسه. فاختبار الاستعداد الدراسى مثلا يطبق على عينة عشوائية ممثلة لمجتمع الطلاب الذين يودون الالتحاق بالكليات، وفى نهاية السنة الجامعية الأولى نوجد متوسط تقديرات هؤلاء الطلاب. ثم نوجد قيمة معامل الارتباط بين درجات الاختبار ومتوسط التقديرات، ونستخدم هذه القيمة فى إيجاد معادلة خط الانحدار الذى نتنبأ بواسطتها بمتوسط تقديرات عينات أخرى من الطلاب لم تلتحق بعد بالكليات بمعلومية اختبار الاستعداد الدراسى الذى يطبق عليهم قبل الالتحاق.

معادلة خط انحدار المتغير (ص) على المتغير (س) :

يمكن إيجاد معادلة خط الانحدار بطرق متعددة لعل أبسطها المعادلة التالية :

$$\hat{ص} = ر \times \frac{ع_{ص}}{ع_{س}} (س - \bar{س}) + \bar{ص} \quad ..... (٣ _ ١١)$$

حيث صَ ترمز إلى قيمة المتغير ص المتنبأ بها بمعلومية قيمة المتغير (س) .

، ر ترمز إلى معامل الارتباط بين المتغيرين (س) ، (ص).

، $ع_س$ ، $ع_ص$ ترمزان إلى الانحراف المعيارى لكل من المتغير (س) ، (ص) على الترتيب.

، $\bar{ص}$ ترمز إلى قيمة متوسط المتغير ص .

ويتطلب تطبيق هذه المعادلة إيجاد قيمة المتوسط والانحراف المعيارى لكل من المتغيرين (س) ، (ص) ، وكذلك قيمة معامل ارتباط بيرسون بينهما. ولتوضيح ذلك نعود إلى جدول (٣ ـ ٧) لإيجاد معادلة انحدار درجات الاختبار (ص) بمعلومية درجات الاختبار (س)، ولذلك نوجد أولاً قيمة كل من $\bar{س}$ ، $\bar{ص}$ ، $ع_س$ ، $ع_ص$ ، ر فنجد أن :

$\bar{س}$ = ١١ ، $\bar{ص}$ = ١٠

، $ع_س$ = ٦,٩٢٨ ، $ع_ص$ = ٦,٠٨٣

، ر = ٠,٧١٩

$$، \hat{ص} = ر \times \frac{ع_ص}{ع_س} (س - \bar{س}) + ص$$

$$= ٠,٧١٩ \times \frac{٦,٠٨٣}{٦,٩٢٨} (س - ١١) + ١٠$$

$$= ٠,٦٣١ (س - ١١) + ١٠$$

$$\hat{ص} = ٠,٦٣١ س + ٣,٠٥٩$$

وهذه هى معادلة انحدار (ص) على (س) التى تستخدم فى التنبؤ بقيم ص إذا علمنا أى قيمة من قيم (س) وذلك بالتعويض عن قيمة س فى هذه المعادلة.

ومن الجدير بالذكر أنه يمكن استخدام برنامج الحاسوب الذى ذكرناه فيما سبق فى إيجاد معادلة انحدار المتغير (ص) على المتغير (س). فقيمة كل من A , B اللذين نحصل عليهما يمكن التعويض بهما فى معادلة الخط المستقيم الذى يمثل معادلة الانحدار وهى :

$$\hat{ص} = B س + A \quad .... \quad (٣ ـ ١٢)$$

حيث B ترمز إلى ميل خط انحدار ص / س

، A ترمز إلى نقطة تقاطع خط الانحدار مع المحور الأفقى.

والمعادلتان (٣ ـ ١١) ، (٣ ـ ١٢) همـا صيغتان مختلفتان ولكن نتائجهما واحدة.

الخطأ المعيارى للتنبؤ:

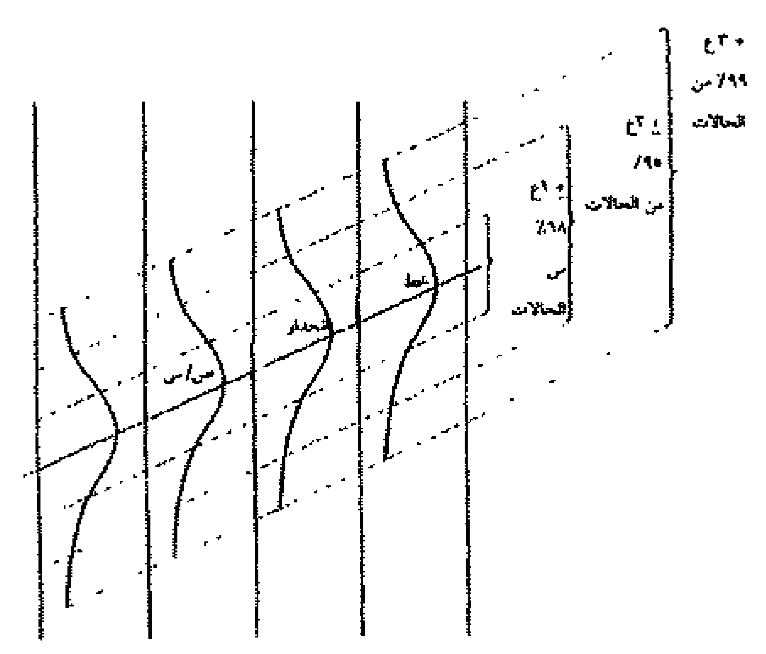
لتقدير أخطاء التنبؤ نستخدم الصيغة التالية التى تعتمد على الانحراف المعيارى للمتغير المتنبأ به (ص)، ومعامل الارتباط بين (س) ، (ص)، ولذلك يسمى الخطأ المعيارى للتنبؤ.

$$\text{الخطأ المعيارى للتنبؤ} = ع \sqrt{١ - ر^{٢}} \quad .... \quad (٣ - ١٢)$$

$$\text{فالخطأ المعيارى للتنبؤ فى المثال السابق} = ٦,٠٨٣ \sqrt{١ - (٠,٧١٩)^{٢}}$$

$$= (٦,٠٨٣)(٠,٦٩٥) = ٤,٢٣$$

ولتفسير مقدار هذا الخطأ نفترض أن أخطاء التنبؤ تتوزع توزيعًا اعتداليًا انحرافه المعيارى ع ، فإذا رسمنا خطوطًا موازية لخط الانحدار على كل من جانبيه على مسافات تساوى قيمة الخطأ المعيارى ٤,٢٣ ومضاعفاتها ، فإنه بالرجوع إلى المساحات تحت المنحنى الاعتدالى المعيارى الموضح بشكل (٣ ـ ١٣) نجد أن حوالى ٦٨ ٪ من الطلاب يقعون بين +١ ، -١ خطأ معيارى ، ٩٥٪ منهم يقعون بين +٢ ، -٢ خطأ معيارى، وهكذا كما هو موضح بشكل (٣ ـ ١٧) التالى :



شكل (٣ ـ ١٧) يوضح توزيع الدرجات المتنبأ بها حول خط انحدار ص على س

أى أن ٦٨ ٪ منهم تنحصر درجاتهم بين ± ٤,٢٣ حول الدرجة صَ المتنبأ بها، ٩٥٪ منهم تنحصر درجاتهم بين (+ ٢ × ٤,٢٣ ، - ٢ × ٤,٢٣)، أى بين ± ٨,٤٦ حول الدرجة صَ.

وبالطبع كلما زاد عدد الأفراد زاد اقتراب عدد القيم التى تنحصر بين الخطين من القيم المتوقعة من التوزيع الاعتدالى. كما أنه كلما ارتفعت قيمة معامل الارتباط كان التنبؤ بقيمة المتغير (ص) أكثر دقة. فمعامل الارتباط ٠,٩٠ مثلا أكثر فائدة فى التنبؤ من المعامل ٠,٢٠ ، ولكن المعامل - ٠,٩٠ يكون مفيدًا فى التنبؤ تمامًا كالمعامل + ٠,٩٠ . أما إذا كانت قيمة معامل الارتباط (-١) أو (+١) فإن جميع النقط تقع على خط الانحدار ، ويكون التنبؤ عندئذ تامًّا.

وفى ختام هذا الفصل نود أن نشير إلى أهمية معامل الارتباط والانحدار الخطى فى تحديد بعض الخصائص السيكومترية المهمة لأدوات القياس وبخاصة الاختبارات مرجعية الجماعة أو المعيار Norm - Referenced Tests ، وهو ما سنتناوله بالتفصيل فى الفصل التالى.

خلاصة

يعتمد القياس التربوى والنفسى على جمع بيانات عن طريق أدوات قياس متنوعة. وهذه البيانات إما أن تكون كمية أو كيفية، وتتطلب البيانات الكمية تبويبا وتصنيفًا وتحليلاً لاستخلاص معلومات مفيدة تتعلق بسمات معينة. ويبدأ هذا التحليل لدرجات الاختبار بتكوين جدول توزيع تكرارى وتمثيله بيانيًا بطرق متعددة، مثل المدرج التكرارى والمضلع التكرارى والمنحنى التكرارى، وكذلك يمكن استخدام شكل الأغصان والأوراق الذى يُعد بمثابة توزيع تكرارى ومدرج تكرارى للدرجات فى آن واحد، ويُوضح الشكل العام لهذا التوزيع، ويُحافظ على جميع الدرجات مرتبة ويمكن حساب ثلاثة مقاييس للنزعة المركزية، هى: المتوسط، والوسيط، والمنوال. ويعتمد اختيار المقياس المناسب على مستوى قياس السمة وطبيعتها، وشكل توزيعها.

ويمكن تحديد مقدار تشتت توزيع الدرجات باستخدام الانحراف المعيارى ونصف المدى الربيعى، كما يمكن استخدام الانحراف المعيارى فى المقارنة بين توزيعين أو أكثر من الدرجات إذا علمنا قيم متوسط هذه التوزيعات، وكذلك يمكن تحديد المركز النسبى للأفراد فى سمة معينة بتحويل الدرجات الخام إلى درجات معيارية أو معيارية مُعدَّلة، أو درجات تائية أو تائية مُعدَّلة أو مئينيات. وجميع هذه الدرجات المحولة تعتمد على الانحراف المعيارى والمتوسط. ويستخدم المنحنى الاعتدالى أيضًا فى إيجاد المئينيات والرتب المئينية المناظرة لدرجات خام معينة.

أما معامل ارتباط بيرسون فيستخدم فى إيجاد العلاقة بين مجموعتين من الدرجات، وتتراوح قيمه بين - ١,٠٠ ، + ١,٠٠ . ويُفيد معامل الارتباط فى وصف درجة العلاقة واتجاهها، كما يُفيد فى التنبؤ بدرجة الفرد فى أحد المتغيرين إذا علمنا درجته فى المتغير الآخر، وذلك باستخدام معادلة الانحدار الخطى البسيط، وتحديد الخطأ المعيارى للتنبؤ الذى يدل على مدى دقة التنبؤ.

الفصل الرابع

الخصائص السيكومترية للاختبارات
أولا : ثبات درجات الاختبارات

* مقدمة
* ثبات درجات الاختبارات
* الثبات والدرجات الحقيقية
* المفهوم الإحصائى لمعامل الثبات
* مؤشر الثبات
* مصادر أخطاء القياس
* أنواع معاملات ثبات درجات الاختبارات
* كيفية تقدير قيم أنواع معاملات الثبات
* معامل التكافؤ
* معامل الاستقرار
* معامل الاستقرار والتكافؤ
* معامل الاتساق الداخلى
* معامل التجانس لكيودر وريتشاردسون
* معامل (α) لكرونباك
* ثبات تقديرات المحكمين
* بعض الطرق الأخرى لتقدير الثبات
* ثبات الفرق بين درجتى اختبارين
* معامل إمكانية التعميم
* تعقيب عام على طرق تقدير الثبات وتفسير نتائجها
* الخطأ المعيارى للقياس
* تفسير الخطأ المعيارى للقياس
* عوامل أخرى تؤثر فى قيم معامل الثبات

مقدمة :

أوضحنا في الفصل السابق كيفية تنظيم ووصف توزيعات درجات الاختبارات والمقاييس التربوية والنفسية باستخدام أساليب إحصائية متعددة لكي يتسنى تفسير الدرجات تفسيراً يفصح عما تتضمنه من معلومات مفيدة تحقق الهدف من عملية القياس. غير أن الاختبارات والمقاييس تعد أدوات قياس وينبغي أن تتوافر فيها شروط معينة لكي تكون صالحة لقياس سمة معينة.

فبعض أدوات القياس والتقويم التربوي والنفسي ربما لا تكون معدة إعداداً جيداً مما يتطلب الحيطة أو الحذر عند استخدامها. وهذا الأمر لا يقتصر على العلوم السلوكية، وإنما يمتد ليشمل أيضاً أدوات القياس في العلوم الطبيعية، فما بالك بميزان أو ترمومتر أو أميتر أو ميكرومتر تعطي قراءات مختلفة إذا تكرر استخدامها في قياس الشيء نفسه في أوقات مختلفة. أليس هذا ممكناً ؟ بل وربما يحدث كثيراً في حياتنا اليومية.

وما بالك إذا كانت لا تقيس الظاهرة التي هدفت لقياسها، وإنما تقيس ظاهرة مختلفة، فعندئذ تكون أداة القياس غير موثوق بها، ويجعل الاعتماد عليها في القياس مضللاً. غير أن مشكلة صلاحية وجودة أدوات القياس تكون أكثر إلحاحاً وتعقيداً في العلوم السلوكية. فالدرجات المستمدة من هذه الأدوات ينبغي أن تتميز بالاتساق ولا تتغير تغيراً جوهرياً من تطبيق إلى آخر على الفرد نفسه ما دامت الظروف لم تتغير تغيرا ملحوظاً. كما ينبغي أن نطمئن إلى أن أداة القياس تقيس بالفعل السمة المحددة لكي نفيد من نتائجها في اتخاذ قرارات صائبة تتعلق بالفرد أو بالجماعة، سواء في الانتقاء أو التصنيف أو التشخيص أو العلاج.

وتشير هذه الخصائص إلى مفهومين من المفاهيم الأساسية التي تتعلق بالاختبارات والمقاييس التربوية والنفسية هما مفهوم ثبات درجات الاختبار Test Reliability، وصدق الاختبار Test Validity. ونظراً لأهمية هذين المفهومين، فإننا سنفرد هذا الفصل لمناقشة مفهوم الثبات في الاختبارات مرجعية الجماعة أو المعيار، ونعرض مفهوم الصدق في الفصل الخامس.

وسوف نتناول الطرق المختلفة لتقدير كل من المفهومين، وكيفية تفسيرهما والعوامل المؤثرة في كليهما ، وعلاقة كل منهما بالآخر.

ثبات درجات الاختبارات :

يُقصد بمفهوم ثبات درجات الاختبارات مدى خلوها من الأخطاء غير المنتظمة التي تشوب القياس، أي مدى قياس الاختبار للمقدار الحقيقي للسمة التي يهدف لقياسها، فدرجات الاختبار تكون ثابتة Reliable إذا كان الاختبار يقيس سمة معينة قياسًا متسقًا في الظروف المتباينة التي قد تؤدي إلى أخطاء القياس. فالثبات بهذا المعنى يعني الاتساق أو الدقة في القياس.

والأخطاء غير المنتظمة التي تؤثر في درجات الاختبارات تكون أخطاء عشوائية يصعب التنبؤ بها من موقف إلى آخر ولذلك تعمل على خفض ثبات الدرجات. وهذه الأخطاء العشوائية ترجع إلى عوامل بعضها يتعلق بالاختبار،مثل عدم وضوح مفرداته، وغموض تعليماته، وعدم تحديد محكات تصحيح مفرداته، والبعض الآخر يتعلق بالظروف البيئية مثل الإضاءة والتهوية والضوضاء وملاءمة غرفة الاختبار، وكذلك عوامل تتعلق بخصائص الأفراد المختبرين،مثل قلة دافعيتهم، وشعورهم بالتعب أو الملل أو القلق، وحالتهم النفسية والصحية بعامة وقت إجراء الاختبار. وكذلك الأخطاء المنتظمة ربما تؤدي إلى تضخيم أو تقليل ثبات درجات الاختبار ولكن بطريقة متسقة، وبالتالي لا تؤثر في ثبات الدرجات.

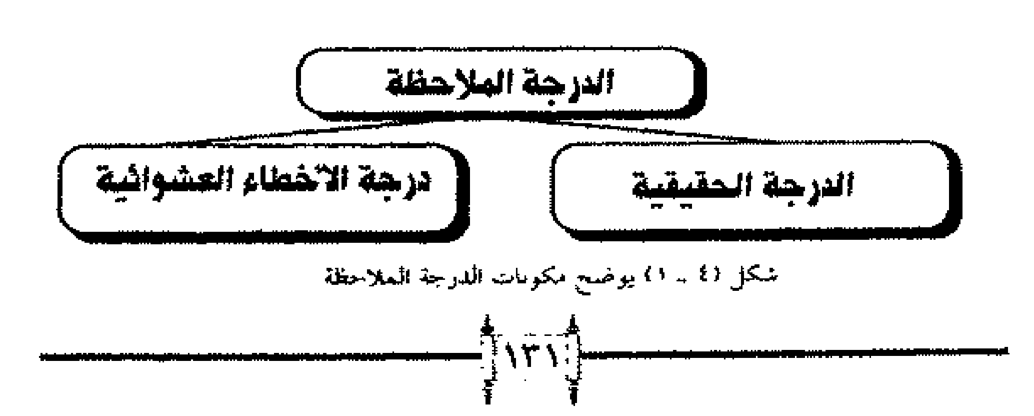
الثبات والدرجات الحقيقية :

إن الدرجة التي يحصل عليها الفرد في مقياس معين تسمى الدرجة الملاحظة Obtained Score ، غير أن هذه الدرجة تكون في كثير من الأحيان مشوبة بأخطاء القياس التي إذا أمكن تحديد مقدارها فإننا نحصل على درجة الخطأ Error Score،وإذا طرحنا درجة الخطأ من الدرجة الملاحظة فإننا نحصل على درجة الفرد خالية من الأخطاء العشوائية، وهذه تسمى الدرجة الحقيقية True Score ، أي أن :

الدرجة الحقيقية = الدرجة الملاحظة − درجة الأخطاء العشوائية

أو أن الدرجة الملاحظة = الدرجة الحقيقية + درجة الأخطاء العشوائية

ويمكن توضيح ذلك بالشكل التخطيطي (٤ ـ ١) التالي:



شكل (٤ ـ ١) يوضح مكونات الدرجة الملاحظة

فالدرجة التى يحصل عليها الفرد فى الاختبار إذن ليست هى الدرجة الحقيقية، وإنما هى الدرجة التى تتكون من الدرجة الحقيقية مضافًا إليها درجة الأخطاء العشوائية كما هو مبين بشكل (٤ ـ ١) . أى أن الدرجة الملاحظة للفرد فى الاختبار لا تساوى درجته الحقيقية إلا إذا كانت درجة الخطأ صفرًا، ولكنها تزيد عن درجته الحقيقية إذا كانت درجة الخطأ موجبة، وتقل عنها إذا كانت درجة الخطأ سالبة .

فإذا لجأ الفرد إلى التخمين فى إجابته عن مفردات الاختبار وتوصَّل إلى الإجابة الصحيحة فإن درجته الملاحظة تزيد عن درجته الحقيقية، أما إذا كان الفرد متعبًا أو أن مفردات الاختبار أو تعليماته غامضة فإن درجته الملاحظة سوف تقل عن درجته الحقيقية.

غير أن هذه الأخطاء الموجبة والسالبة يلاشى بعضها البعض الآخر على المدى البعيد استنادًا إلى قوانين الصدفة، وبذلك يصبح متوسط درجات الخطأ صفرًا. ولذلك فإن متوسط الدرجات الملاحظة لمجموعة الأفراد الذين يطبق عليهم الاختبار يعد أفضل تقدير لمتوسط درجاتهم الحقيقية فى الاختبار.

كما أن تباين الدرجات الملاحظة يساوى مجموع تباين كل من الدرجات الحقيقية وتباين الخطأ (لاحظ أن مقدار التباين = مربع مقدار الانحراف المعيارى(كما سبق أن أوضحنا فى الفصل الثالث)، فإذا كانت الدرجات الملاحظة خالية من الخطأ فإن مقدار تباين الخطأ يساوى صفر، ويصبح مقدار تباين الدرجات الملاحظة مساويًا مقدار تباين الدرجات الحقيقية، أى تكون الدرجات متسقة اتساقًا تامًا، وتكون قيمة معامل الثبات واحدًا صحيحًا .

أما إذا كانت درجات الاختبار غير متسقة على الإطلاق، فمعنى ذلك أن الدرجات لا تمثل سوى أخطاء القياس، أى أن الفروق الملاحظة بين الأفراد فى هذه الحالة لا يكون لها أى علاقة بالفروق الحقيقية بينهم، وتكون قيمة معامل الثبات صفرًا.

لذلك فإن الانحراف المعيارى لدرجات الخطأ يسمى الخطأ المعيارى للقياس Standard Error of Measurement . فإذا كان مقدار هذا الخطأ المعيارى صفرًا تكون درجات الاختبار متسقة اتساقًا تامًا. ولكننا نلاحظ عادة فى القياس التربوى والنفسى أن هناك قدرًا معينًا من الخطأ المعيارى مما يؤدى إلى خفض ثبات درجات الاختبارات بقيم متفاوتة بحسب مقدار هذا الخطأ. وسوف نناقش مفهوم الخطأ المعيارى للقياس فيما بعد .

المفهوم الإحصائي لمعامل الثبات :

على الرغم من أن الدرجات الحقيقية ودرجات الأخطاء العشوائية تعد من المفاهيم الافتراضية، إلا أنها تفيد بدرجة كبيرة في التوصل إلى المفهوم الإحصائي للثبات، فقد تبين لنا أنه إذا كانت درجات الخطأ غير مرتبطة بالدرجات الحقيقية فإن تباين الدرجات الملاحظة = تباين الدرجات الحقيقية + تباين الخطأ.

أي أن $ع_{\text{م}}^{2} = ع_{\text{ح}}^{2} + ع_{\text{خ}}^{2}$ ......(٤ ـ ١)

وإذا قسمنا كلا من طرفي المعادلة (٤ ـ ١) على $ع_{\text{م}}^{2}$ نحصل على :

$$1 = \frac{ع_{\text{ح}}^{2}}{ع_{\text{م}}^{2}} + \frac{ع_{\text{خ}}^{2}}{ع_{\text{م}}^{2}} ،$$

أي أن : $\frac{ع_{\text{ح}}^{2}}{ع_{\text{م}}^{2}} = 1 - \frac{ع_{\text{خ}}^{2}}{ع_{\text{م}}^{2}}$ .... (٤ ـ ٢)

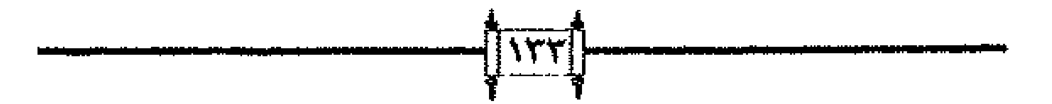
والكسر $\frac{ع_{\text{ح}}^{2}}{ع_{\text{م}}^{2}}$ يدل على نسبة التباين الحقيقي إلى التباين الملاحظ.

أما الكسر $\frac{ع_{\text{خ}}^{2}}{ع_{\text{م}}^{2}}$ فيدل على نسبة تباين الخطأ إلى التباين الملاحظ.

ولذلك فإن الكسر الأول يُعبر عن معامل ثبات درجات الاختبار Coefficient of Reliability . أي أن معامل الثبات من المنظور الإحصائي يمكن اعتباره نسبة تباين الدرجات الحقيقية إلى تباين الدرجات الملاحظة كالتالي:

$$\text{معامل الثبات } ر_{11} = \frac{\text{تباين الدرجات الحقيقية}}{\text{تباين الدرجات الملاحظة}} \quad \ldots\ldots (٤ـ٣)$$

وهذا يعني أن معامل الثبات هو ذلك الجزء من تباين الدرجات الملاحظة الذي يُعزى إلى تباين الدرجات الحقيقية. وتتراوح قيم هذا المعامل بين صفر، ١,٠ ، ولذلك



يعبر عن قيمته بكسر عشرى مثل معامل الارتباط. فكلما اقتربت قيمته من الواحد الصحيح دل ذلك على قدر أكبر من ثبات الدرجات، وبالتالى تزداد الثقة فى دقة هذه الدرجات واتساقها والاعتماد عليها. ويمكن توضيح ذلك بشكل (٤ ـ ٢) التالى:

معامل ثبات منخفض

تباين الدرجات الحقيقية	تباين الخطأ

معامل ثبات مرتفع

تباين الدرجات الحقيقية	تباين الخطأ

تباين الدرجات الملاحظة

شكل (٤ ـ ٢) يوضح زيادة قيمة الثبات بزيادة القدر النسبى

لتباين الدرجات الحقيقية

العلاقة بين الدرجات الحقيقية والدرجات الملاحظة (مؤشر الثبات):

لعل من المفيد أن نوضح العلاقة بين الدرجات الحقيقية لمجموعة من الأفراد ودرجاتهم الملاحظة فى اختبار ما فى ضوء المفهوم الإحصائى لمعامل الثبات.

وهذا يتطلب العودة إلى الفصل الثالث لمراجعة مفهوم معامل الارتباط، مع ملاحظة أنه يمكن كتابة الصيغة (٣ ـ ١٠) بطريقة مختصرة كالتالى:

$$\text{معامل الارتباط ر} = \frac{\text{مجـ س ص}}{\text{ن} \times \text{ع}_{\text{س}} \times \text{ع}_{\text{ص}}} \quad \text{...........(٤ ـ ٤)}$$

حيث ترمز (س) ، (ص) إلى انحراف كل منهما عن متوسطها.

وترمز ($\text{ع}_{\text{س}}$) ، ($\text{ع}_{\text{ص}}$) إلى الانحراف المعيارى لكل من (س) ، (ص) .

وترمز (ن) إلى عدد أزواج الدرجات .

فإذا رمزنا لانحرافات كل من الدرجات الحقيقية، والدرجات الملاحظة، ودرجات الخطأ عن متوسط توزيعاتها المناظرة بالرموز (ح) ، (م) ، (خ) على الترتيب، وتطبيق الصيغة (٤ـ٤) ، فإن معامل الارتباط بين الدرجات الحقيقية ودرجات الخطأ يكون كالتالى :

$$ر_{حم} = \frac{\text{مجـ}\,(ح \times م)}{ن \times ع_{ح}} \quad \ldots\ldots\ldots\ldots\ldots (٤ـ٤)$$

$$= \frac{\text{مجـ}\,ح\,(ح + خ)}{ن \times ع_{ح} \times ع_{م}} \qquad \text{(وذلك لأن م = ح + خ)}$$

$$= \frac{\text{مجـ}\,(ح^{٢} + ح\,خ)}{ن \times ع_{ح} \times ع_{م}} = \frac{\text{مجـ}\,ح^{٢}}{ن\,ع_{ح}\,ع_{م}} + \frac{\text{مجـ}\,ح\,خ}{ن \times ع_{ح} \times ع_{م}} \quad \ldots (٤-٥)$$

ولكن $\frac{\text{مجـ}\,ح^{٢}}{ن}$ = تباين الدرجات الحقيقية $ع_{ح}^{٢}$

، مجـ ح خ = مجموع حواصل ضرب الدرجات الحقيقية فى درجة الخطأ، ونظراً لأنهما غير مرتبطين كما سبق أن ذكرنا ، فإن هذا المجموع = صفر .

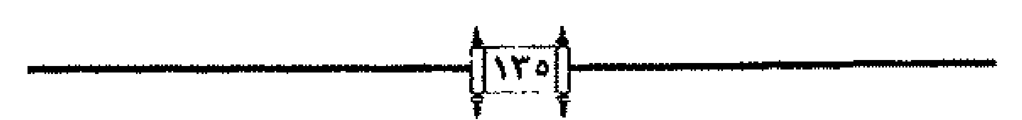
أى أن : مجـ ح خ = صفر ، وبالتالى ينعدم كسر النسبة الثانية .

وبالتعويض فى (٤ ـ ٥) نجد أن :

$$ر_{حم} = \frac{ع_{ح}^{٢}}{ع_{ح} \times ع_{م}} = \frac{ع_{ح}}{ع_{م}} \quad \ldots\ldots\ldots (٤ـ٦)$$

ولكن سبق توضيح أن : معامل الثبات $ر_{مم} = \frac{ع_{ح}^{٢}}{ع_{م}^{٢}}$

فبالتعويض فى (٤ ـ ٦) نجد أن :



$ر_{م م} = ر^2_{ح م}$ أى أن $ر_{ح م} = \sqrt{ر_{م م}}$ .... (٤ ـ ٧)

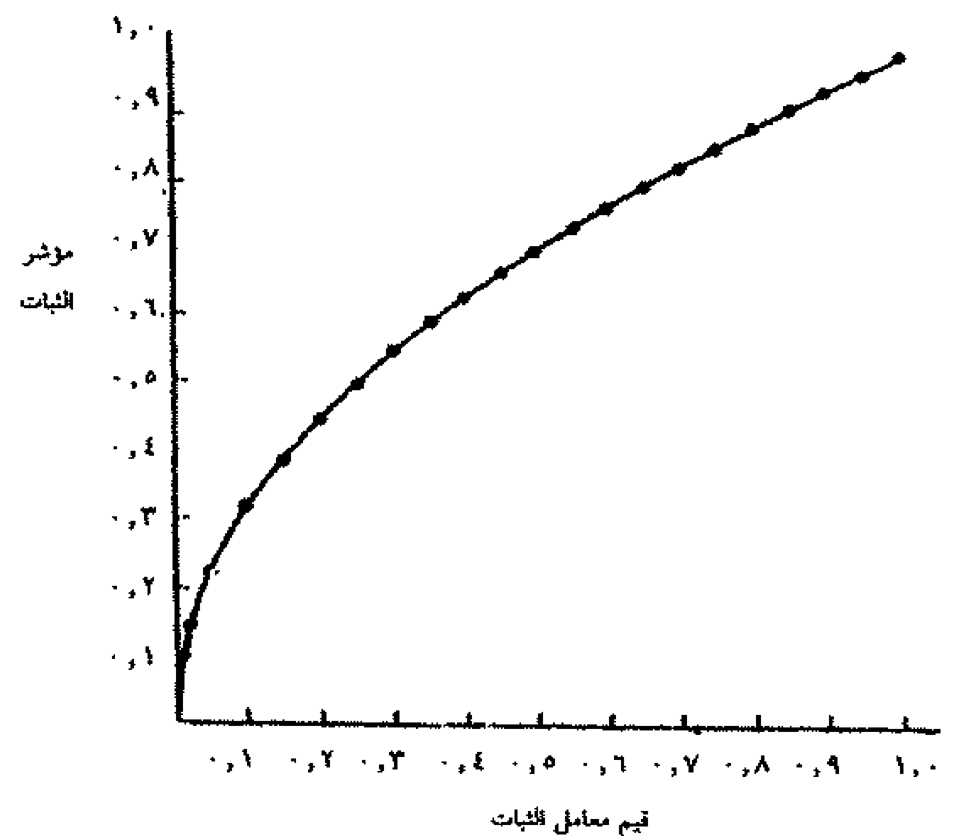
وهذا يعنى أنه يمكن تقدير معامل الارتباط بين الدرجات الحقيقية والدرجات الملاحظة لمجموعة من الأفراد فى اختبار ما بإيجاد الجذر التربيعى لمعامل ثبات درجات هذا الاختبار. ويسمى معامل الارتباط عندئذ « مؤشر الثبات » Reliability Index، فإذا كانت قيمة معامل ثبات درجات اختبار ٠,٨٠ مثلاً فإن القيمة التقديرية لمعامل الارتباط بين الدرجات الحقيقية والدرجات الملاحظة فى الاختبار تساوى $\sqrt{٠,٨٠}$ = ٠,٨٩ وهذه هى قيمة مؤشر الثبات.

ونظرًا لأن معامل الثبات يساوى مربع مؤشر الثبات أى مربع معامل الارتباط بين الدرجات الحقيقية والدرجات الملاحظة، فإن معامل الثبات يعد بمثابة معامل التحديد الذى أوضحناه عند مناقشتنا لكيفية تفسير معامل الارتباط. فمعامل الثبات إذن يحدد نسبة تباين الدرجات الملاحظة للفرد التى تعزى إلى التباين الحقيقى للدرجات فيما يقيسه الاختبار. وإذا طرحنا قيمته من الواحد الصحيح، فإننا نحصل على نسبة التباين فى الدرجات الملاحظة التى تعزى إلى الأخطاء العشوائية.

وتجدر الإشارة هنا إلى أن قيمة معامل الثبات هى القيمة التى يذكرها الباحثون ومعدو الاختبارات والمقاييس عادة فى أدلة الاختبارات. فمعامل الثبات يدل على العلاقة بين مجموعتين من الدرجات الملاحظة (المعلومة)، مثل درجات اختبارين متكافئين يقيسان السمة نفسها، بينما يدل مؤشر الثبات على العلاقة بين الدرجات الحقيقية (غير المعلومة) والدرجات الملاحظة (المعلومة) . لذلك فإن مؤشر الثبات يعطى قيمة تقديرية لهذه العلاقة ، وعلى الرغم من ذلك فإن كليهما يقدم نفس المعلومات، ولكن بطريقتين مختلفتين.

وقد أوضح جاليكسان Gulliksan أن مقدار صدق الاختبار ينبغى أن يكون دائمًا أقل من قيمة مؤشر الثبات، أى أقل من $\sqrt{ر_{م م}}$ كما سنرى عند مناقشتنا لمفهوم الصدق. فمن الناحية النظرية لا ينبغى أن ترتبط درجات اختبار معين بأى متغير آخر بدرجة أكبر من ارتباطها بالمتغير الذى يقيسه الاختبار، أى ارتباط الاختبار بنفسه.

ويمكن توضيح العلاقة بين قيم كل من معامل الثبات، ومؤشر الثبات بيانيًا بتمثيل قيم معامل الثبات التى تتراوح بين صفر ، ١,٠ على المحور الأفقى ، وقيم مؤشر الثبات والتى تتراوح أيضًا بين صفر ، ١,٠ على المحور الرأسى . كما هو موضح بالشكل (٤ ـ ٣) التالى :



شكل (٤ ـ ٣) يوضح العلاقة المنحنية بين معامل الثبات ومؤشر الثبات

وبالنظر إلى شكل (٤ ـ ٣) يتضح أن قيم مؤشر الثبات تكون أكبر من قيم معامل الثبات المناظرة لها فيما عدا عندما تكون قيمة معامل الثبات صفرًا أو واحدًا صحيحًا. فمؤشر الثبات يساوى ٠,٨٩ إذا كان معامل الثبات ٠,٨٠ ، كما سبق أن رأينا ، ويصبح مؤشر الثبات مساويًا ٠,٩٥ إذا كان معامل الثبات ٠,٩٠ وهكذا.

ويمكن الحصول على قيم مؤشر الثبات برسم خط رأسي من النقطة الممثلة لقيمة معامل الثبات ، وعند التقاء هذا الخط بالمنحنى نرسم خطًا أفقيًا يلاقى المحور الرأسى عند نقطة تمثل قيمة مؤشر الثبات.

مصادر أخطاء القياس :

ذكرنا أن هناك نوعين من أخطاء القيـاس أحدهما الأخطاء المنتظمة Systematic Errors،والأخرى الأخطـاء غير المنـتظمة أو العشـوائية Unsystematic Errors التى ترجع إلى عوامل الصدفة Chance Factors .

وهذه الأخطاء العشوائيـة هي تلك الأخطاء التى لا ترتبط بأداء الفرد المـستقبلى الذى نود الاستـدلال عليه، وبالتـالى فهى لا ترتبـط بدرجة الفرد الحـقيقيـة أو بدرجة الأخطاء العشوائية إذا ما أعيد تطبيق الاختبار نفسه أو صيغة مكافئة له على الفرد نفسه.

فدرجة الفرد الحقـيقية فى الاختبار هى تلك الدرجة الناجمـة عن جميع العوامل المنتظمة التى نقررها بما فى ذلك الأخطاء المنتـظمة التى ربما تسهم بدرجة منتظمة فى تباين درجات الاختبار وتكون جزءاً من تباين الدرجـة الحقيقية. ولتوضيح ذلك نفترض أننا اختبرنا مجموعة من الأفراد بمفردات اختـبارية مختلفة فعندئذ تسهم أخطاء الصدفة فى تباين درجاتهم الحقيقية نظراً لاختـلاف عينة المفردات، أما إذا اختبرنا جميع الأفراد بعينـة واحدة من المفردات، فـإننا بذلك نستبـعد مثل هذه الأخطاء عـلى الرغم من أن اختلاف طبيعة مفردات هذه العينة بالنسبة لكل فرد يمكـن أن تؤدى إلى تباين إجاباتهم عنها، ولكن بـطريقة منتظمة. وكذلك إذا أضـفنا لكل طالب درجتين عـلى درجته فى الاختبار، فإن هذه الإضافة تعد من الأخطاء المنتظمة التى لا تؤثر فى قيمة معامل ثبات الاختبار.

ولذلك فإن هذه الأخطـاء المنتظمة أو الثابتـة لا تهمنا كثيـراً فى هذا الشأن وإن كانت تؤثر بلا شك فى دقة البحث العلمى.

ونظراً لأن هذا النوع من الأخطاء له علاقة بصدق الاختبار فسوف نرجئ مناقشته للجزء الخاص بالصدق فى الفصل الخامس.

غير أن الذى يجـب مراعاته جيداً هو الأخطاء غـير المنتظمة أو العشـوائية،حيث إنها تؤثر بدرجات متفاوتة فى معامل الثبات كما سبق أن أوضحنا.

وهنا ربما يتساءل القارئ عن مصادر هذه الأخطاء العشوائية من أجل التحكم فيها أثناء إعداد الاختبار وعند تطبيقه وتصـحيح مفرداته، وذلك لجعل تأثيرها فى الدرجات الحقيـقية أقل ما يمكـن، وبذلك تقترب الدرجـات الملاحظة للأفراد فى الاخـتبار من درجاتهم الحقيقية ويزيد تباين الدرجات الحقيقية إلى أقصى حد ممكن، وبالتالى تكون قيم معامل الثبات أكبر ما يمكن.

ونظراً لأن الأخـطاء العشـوائية الـتى تؤثر فى مـوقف اخـتـبارى معـين تخـتلف باختلاف مجـموعة الأفراد المختبرين، ونـوع الاختبار وعينة مفرداته، وظـروف تطبيقه



وتصحيح مفرداته، فإن مصادر هذه الأخطاء تكون كثيرة ومتعددة. غير أننا سوف نقتصر فيما يلي على مناقشة بعض أهم هذه المصادر التي تؤثر في القياس التربوي والنفسي، ويمكن تصنيفها في ثلاثة مصادر أساسية كالتالي :

أولا : مصادر تتعلق بأداة القياس:

تعد أدوات القياس في بعض الأحيان مصدراً من مصادر الأخطاء العشوائية. فعلى الرغم من أنه ينبغي أن يراعى في تصميم هذه الأدوات جعل هذه الأخطاء أقل ما يمكن، إلا أننا نلاحظ أن مفردات الاختبار التربوي أو النفسي قد تكون غاية في الصعوبة، أو تحتمل التخمين أو أن تكون صياغتها غامضة أو مربكة، أو أن تكون تعليمات الإجابة عن الاختبار غير محددة، أو تشجع على التخمين كما في مفردات الصواب أو الخطأ ، أو الاختيار من متعدد، وما إلى ذلك من العوامل التي تقلل من ثبات درجات الاختبار.

غير أن المصدر الأساسي للأخطاء العشوائية المتعلقة بالاختبار تتعلق بأسلوب معاينة المفردات للنطاق السلوكي الذي يقيسه الاختبار مما يترتب عليه أن يكون عدد المفردات قليلاً أو أن المفردات لا تمثل مكونات السمة المقاسة تمثيلاً كافيًا.

ولكي يتضح الإطار النظري لهذا النوع من الأخطاء ينبغي أن نلقي بعض الضوء على أهم النماذج التي تتناولها النظرية الكلاسيكية للاختبارات Classical Test Theory ، فهذه النظرية تتعلق بالاختبارات والمقاييس مرجعية الجماعة أو المعيار Norm - Referenced Tests التي تهتم بقياس الفروق الفردية والكشف عنها، لذلك تفترض أن الأفراد المختبرين تتوزع درجاتهم توزيعًا اعتداليًا على متصل السمة أو القدرة التي يقيسها الاختبار. ويتحقق ذلك باستخدام أساليب انتقاء المفردات الاختبارية التي تجعل توزيع الدرجات خاضعًا لهذا التوزيع الاعتدالي Normal Distribution. ويندرج تحت هذه النظرية نموذجين أحدهما يسمى نموذج معاينة النطاق The Domain Sampling Model ، والآخر نموذج الاختبارات المتوازية The Parallel Tests Model، وكلا النموذجين يهتمان بالأخطاء العشوائية للقياس.

(١) **نموذج معاينة النطاق** : يعتبر هذا النموذج أن أي اختبار أو قياس يشتمل على عينة عشوائية من المفردات المسحوبة من نطاق أو مجتمع افتراضي من المفردات، مثل اختبار عمليات الجمع البسيط الذي يفترض أنه يشتمل على عينة عشوائية من الأعداد ذات الرقمين التي تنحصر بين ١٠ ، ١٠٠ مثلا. وعلى الرغم من أن هذا الافتراض ربما لا يكون واقعيًا، إلا أن النموذج يمكن أن يؤدي إلى تقديرات دقيقة للدرجات الحقيقية نظرًا لأن مصمم الاختبار يكتب عادة عددًا كبيرًا من مختلف أنواع المفردات في اختبار معين ويعتبرها كما لو كانت عينة عشوائية من المفردات. فالهدف الرئيسي



من بناء الاختبارات والمقاييس فى إطار هذا النموذج هو تقدير القياسات التى يمكن الحصول عليها إذا اختُبر الفرد بجميع المفردات التى يشتمل عليها النطاق أو المجتمع (أى جميع الأعداد ذات الرقمين مثلا) . والدرجة التى يمكن أن يحصل عليها الفرد فى النطاق الشامل للمفردات Items Universe تسمى الدرجة الحقيقية True Score أو درجة النطاق الشامل Universe Score ،فإذا ارتبطت عينة المفردات التى يُختبر بها الأفراد ارتباطًا مرتفعًا بالدرجة الحقيقية، يقال: إن عينة مفردات الاختبار تتميز درجاتها بالثبات. لذلك يختص هذا النموذج بالاختبارات المتوازية عشوائيًا Randomly Parallel Tests حيث تختلف درجاتها الكلية إلى حد ما فى المتوسط والانحراف المعيارى والارتباط فيما بينها نظرًا إلى الأخطاء العشوائية فى معاينات المفردات التى يشتمل عليها الاختبار (Lord & Novick , 1950 ; Gulliksen ، 1950).

(٢) نموذج الاختبارات المتوازية :

يفترض هذا النموذج التوازى الفعلى للاختبارات وليس توازيها على أساس المعاينات العشوائية. أى يفترض تساوى متوسطات درجات اختبارات معينة وتساوى انحرافاتها المعيارية، وارتباط درجاتها الملاحظة بالدرجات الحقيقية ارتباطًا متساويًا فى قيمه. وتباين درجات أى من هذه الاختبارات التى لا يعزى إلى الدرجات الحقيقية إنما يرجع إلى محض الصدفة أو الأخطاء العشوائية.

وعلى الرغم من هذا الاختلاف فى فرضيات كل من النموذجين، إلا أنهما يتوصلان إلى نتائج متماثلة فيما يتعلق بالأخطاء العشوائية للقياس.

غير أن نانلى Nunnally يرى أنه يصعب التحقق الإمبيريقى المباشر من إحدى فرضيات النموذج الثانى المتعلقة بقيم الارتباط بين الدرجات الملاحظة والدرجات الحقيقية.

وفى ضوء هذين النموذجين نلاحظ أن المصدر الأساسى لعدم اتساق درجات الاختبارات والمقاييس الذى يرجع إلى الاختبار أو المقياس نفسه يتعلق بأسلوب معاينة نطاق المفردات التى يتكون منها الاختبار.

فإذا كان للاختبار صيغ متوازية أو متكافئة فإنه ينبغى أن تتحقق فيها شروط أى من النموذجين. ونظرًا لصعوبة الانتقاء العشوائى لمفردات الاختبار من نطاق شامل لهذه المفردات، فإنه يمكن المزاوجة بين المفردات فى الصيغ المتعددة للاختبار من حيث مكونات السمة التى تقيسها، ومحتوى المفردات ودرجة صعوبتها، وتماثل توزيع درجاتها.

وهذا بالطبع يتطلب العناية بتصميم الاختبار وبناء مفرداته وتحليل نتائجها ، وإذا لم يحظ الاختبار بمثل هذه العناية، فإن درجات الأفراد سوف تتباين من صيغة إلى أخرى للاختبار نفسه ، مما يجعل الدرجات مشوبة بأخطاء القياس التى تعزى إلى أداة القياس

ثانياً : مصادر تتعلق بإجراءات تطبيق الاختبار وتصحيحه:

إن الاختبارات والمقاييس التربوية والنفسية يجب أن تطبق على الأفراد فى ظروف مقننة. فقد سبق أن أوضحنا فى الفصل الأول أن التقنين مكونة أساسية من مكونات تعريف أداة القياس، وناقشنا عناصر هذه المكونة. وإجراءات تطبيق الاختبار وتصحيحه تعد مصدراً من مصادر الأخطاء العشوائية فى القياس، غير أن هذا المصدر يمكن ضبطه أو التحكم فيه لأنه يتعلق بالبيئة الفيزيائية المحيطة بالفرد أثناء تطبيق الاختبار. فالإضاءة الجيدة لغرفة الاختبار، وحسن تهويتها، ومنع الضوضاء حولها، وتوزيع المقاعد بها وغير ذلك من العوامل الفيزيائية التى ربما تؤثر فى إجابات الأفراد وبالتالى فى درجاتهم الملاحظة ينبغى أن يتم ضبطها قبل بدء العملية الاختبارية.

كما ينبغى أن تكون تعليمات الاختبار واضحة ومفهومة وبخاصة فى الاختبارات الفردية أو التى تطبق على الأطفال أو الأميين، حيث يفضل أن تكون التعليمات شفوية ومناسبة للمستوى العمرى للأطفال. كما يجب أن يقاس زمن الإجابة عن الاختبارات التى تتطلب السرعة قياساً دقيقاً، وأن تكون طباعة الاختبار واضحة، وأوراق الإجابة منظمة بطريقة تسمح للفرد بتحديد إجابته بيسر وسهولة. وهذا يتطلب أن تشتمل أدلة الاختبارات على إشارات تفصيلية لمن يستخدم الاختبار بحيث يمكنه اتباعها بدقة، على أن تتضمن هذه الإشارات كيفية التعامل فى الظروف غير المتوقعة التى يمكن أن تطرأ أثناء تطبيق الاختبار.

وكذلك تشتمل على كيفية تصحيح الاختبار بطريقة موضوعية لتقليل أثر الأخطاء الناجمة عن اختلاف التقديرات وبخاصة إذا كانت مفردات الاختبار تتطلب إجابات مفتوحة كما فى أسئلة المقال أو بعض مقاييس الشخصية والاتجاهات، وكذلك كيفية رصد الدرجات، والدقة فى العمليات الحسابية التى ربما يقوم بها الفاحص للتوصل إلى الدرجة الكلية فى الاختبار وبخاصة إذا كانت تجرى بطريقة يدوية.

وينبغى ملاحظة أن عوامل التحيز فى أى من هذه الإجراءات التى ربما تكون متسقة تؤثر فى التباين الحقيقى للدرجات ولكنها تؤدى إلى خفض صدق الاختبار. فالمعلم الذى يقوم بتقدير درجات اختبار مقالى فى مادة دراسية معينة ربما يتأثر بعوامل لا تتعلق بإجابات الأسئلة ، مثل أخطاء النحو والهجاء، أو جودة الخط، أو فكرته

الذاتية عن الطالب مما يؤثر في الدرجة التي يحصل عليها الطالب، وبالتالي تؤثر في ثبات درجات الاختبار. ولعل الاتفاق على محكات واضحة قبل تقدير درجات الاختبار يسهم في استبعاد مثل هذه العوامل الدخيلة.

ثالثاً: مصادر تتعلق بالأفراد المختبرين:

على الرغم من أن هناك خصائص معينة للأفراد المختبرين تكون متسقة أثناء أدائهم الاختبارات، مثل الخبرة السابقة لدى الفرد ومستوى قدرته، حيث تسهم في التباين الحقيقي للدرجات، إلا أن هناك خصائص أخرى قد تبدو متسقة وتسهم أيضاً في التباين الحقيقي للدرجات ولكنها غير مرغوبة، مثل تذكر الإجابات عن مفردات الاختبار نتيجة الخبرة السابقة أو إعادة تطبيق صيغة مكافئة للاختبار، أو نتيجة مرض مزمن لدى الفرد أو انخفاض دافعيته، أو اتجاهه السلبي، أو تشتت انتباهه بصورة دائمة. وهذا المصدر من مصادر الخطأ يصعب التحكم فيه؛ لأنه يتعلق بالفرد نفسه. فالاختبارات التحصيلية أو اختبارات الذكاء تتطلب من الفرد بذل أقصى جهده للحصول على درجة مرتفعة، غير أن هناك أفراداً يتسمون في معظم الأحيان بانخفاض دافعيتهم للتحصيل أو اتجاههم السلبي نحو الاختبارات، فعندئذ تنخفض درجاتهم في الاختبار نتيجة لذلك. وكذلك في مقاييس الشخصية التي تستخدم بغرض انتقاء أصلح الأفراد للالتحاق بوظيفة أو مهنة معينة، ربما يحاول الفرد إعطاء انطباع جيد للمهتمين بنتائج القياس، وذلك بأن يجيب عن المقياس كما لو كان يتطلب بذل أقصى جهده للحصول على درجة مرتفعة، وبذلك لا يعد مقياساً للأداء المميز الذي يجيب فيه الفرد كما يسلك عادة في حياته اليومية، وإنما قد يجيب إجابات تبدو مرغوبة اجتماعيا. ومع هذا فإن هذه الخصائص المتسقة ربما تكون مرتبطة بقدرات الفرد أو سماته.

غير أن ما يهمنا هو تلك الخصائص التي تسهم في أخطاء القياس وينبغي العمل على تقليل أثرها بقدر الإمكان. وتتعلق هذه الخصائص بالظروف الطارئة أو الوقتية للأفراد مثل التعب، والانخفاض النسبي للدافعية، وتقلب الحالة المزاجية، وقلة الاكتراث بالاختبار، وما شابه ذلك. وعلى الرغم من صعوبة التحكم في هذه المتغيرات أيضاً إلا أن معامل الثبات يجب أن يعكس الأخطاء الناجمة عنها فقط.

والخلاصة أنه ينبغي أخذ جميع هذه المصادر بعين الاعتبار عند تقدير ثبات درجات الاختبار وفي تفسير نتائجه، كما ينبغي مراعاة المصادر الأخرى المتعددة التي قد تؤثر في دقة الدرجات واتساقها.

أنواع معاملات ثبات درجات الاختبارات :

أوضحنا في الجزء السابق المفهوم الإحصائي لمعامل الثبات ومصادر الأخطاء التي يمكن أن تؤثر في قيم هذا المعامل، وتبين لنا أن هناك مصادر متعددة للأخطاء غير المنتظمة أو العشوائية التي ينبغي التقليل من أثرها بقدر الإمكان لكي تعبر الدرجات الملاحظة التي يحصل عليها الأفراد في الاختبارات عن السمات التي تقيسها هذه الاختبارات ممثلة في درجاتهم الحقيقية. ونظراً لأنه يصعب إلغاء تأثير هذه الأخطاء العشوائية في القياس التربوي والنفسي، فإنه يمكن تقدير أثرها في الدرجات الملاحظة بحيث نستطيع معرفة مدى إسهامها في اختلاف هذه الدرجات عن الدرجات الحقيقية، وبالتالي يمكن تقدير قيم معامل ثبات الدرجات الملاحظة.

ولذلك فإن جميع الطرق التي تستخدم في تقدير هذه القيم تعتمد على الدرجات الملاحظة، حيث إن الدرجات الحقيقية تكون غير معلومة.

وتختلف هذه الطرق باختلاف مصدر الأخطاء العشوائية التي نود تقدير تأثيرها في الدرجات الملاحظة، وهذا يعتمد على طبيعة الاختبار وفيم تستخدم نتائجه .

ويتعلق هذا التقدير بمعرفة مدى اتساق درجات الاختبار في قياسات متكررة لنفس الفرد أو مجموعة من الأفراد. ويحدث ذلك عادة في القياس الفيزيائي لتقدير دقة أداة القياس أو عمليات القياس حيث يعاد قياس الظاهرة عدة مرات للتحقق من ذلك.

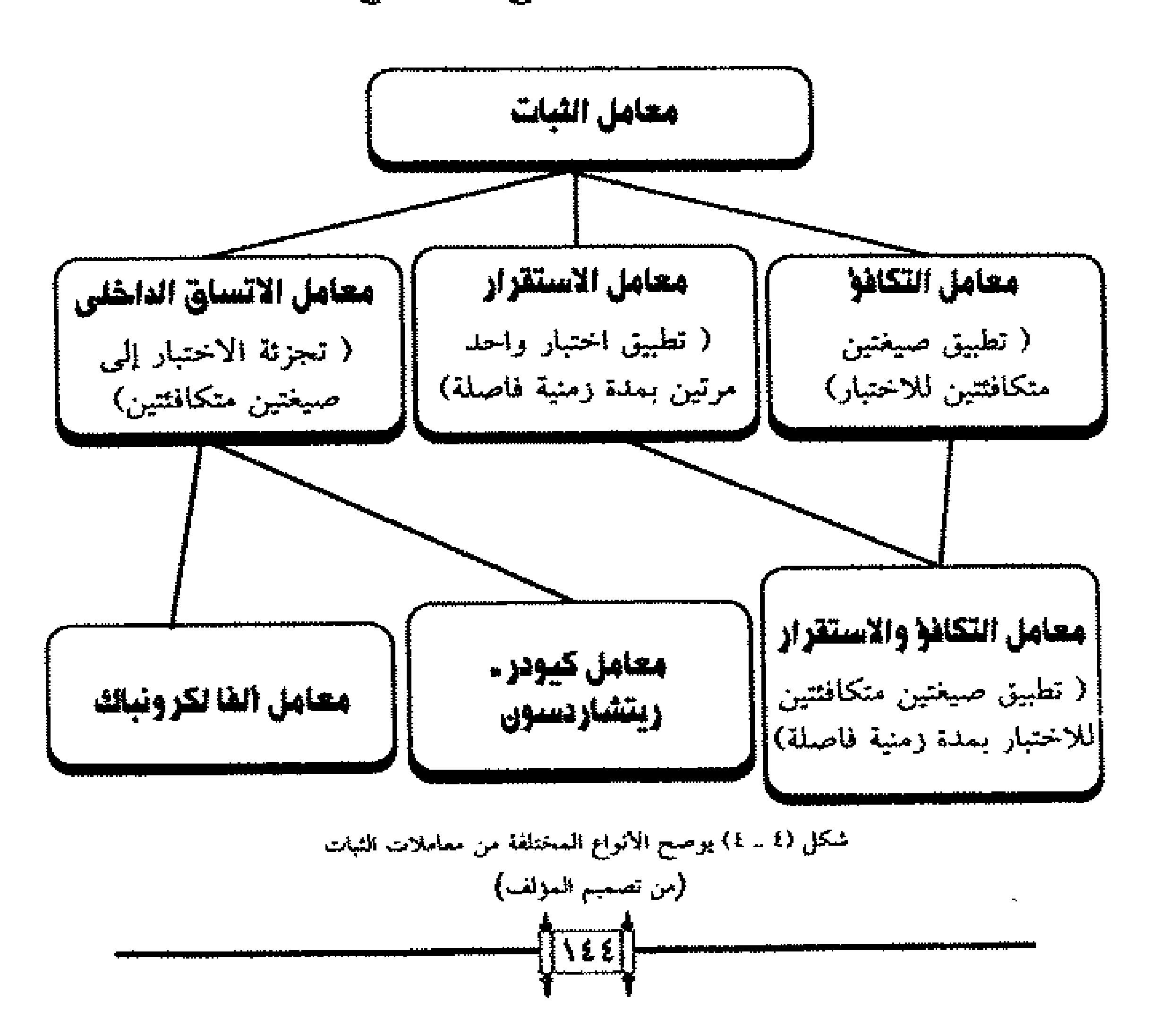
غير أننا في القياس التربوي والنفسي لا نستطيع تكرار عملية القياس على الفرد أو مجموعة الأفراد عدة مرات نظراً لأن هذه العملية تؤدي إلى تغيرات في سلوك الفرد المختبر.

ولذلك فإننا نعتمد على تكرار القياس مرتين على الأكثر للفرد نفسه أو مجموعة الأفراد، وعندئذ نحصل على درجتين لكل فرد. ويكون لدينا مجموعتين من أزواج الدرجات المناظرة لمجموعة الأفراد. ويمكن الحصول على تقدير لمتوسط اتساق درجات المجموعة في الاختبار بإيجاد قيمة معامل ارتباط بيرسون بين أزواج الدرجات، وهذا يعد تقديراً لمعامل ثبات درجات الاختبار.

وقد ذكرنا فيما سبق أن معامل الثبات هو النسبة بين تباين الدرجات الحقيقية وتباين الدرجات الملاحظة، وعند تكرار القياس باستخدام صيغتين متكافئتين للاختبار نفسه على مجموعة واحدة من الأفراد فإن هذه النسبة تساوي قيمة معامل الارتباط بين درجات الاختبارين المتكافئين، وذلك للتداخل التام بين تباين الدرجات الحقيقية لكل منهما، ويساوي النسبة بين تباين الدرجات الحقيقية والتباين الكلي لدرجات كل من الاختبارين.

وتشير أدبيات القياس والتقويم التربوى والنفسى إلى طرق متعددة لتقدير قيم معامل الثبات، بعضها يعتمد على تطبيق صيغتين متكافئتين للاختبار لتقدير تكافؤ Equivalence أو اتساق Consistency درجات الصيغتين، وبعضها يعتمد على المدة الزمنية الفاصلة بين تطبيق اختبار واحد مرتين على مجموعة الأفراد نفسها لتقدير استقرار Stability درجات الاختبار، وبعضها يعتمد على ضم الطريقتين السابقتين معًا بحيث تطبق صيغتان متكافئتان للاختبار بمدة زمنية فاصلة لتقدير معامل الاستقرار والتكافؤ Stability and Equivalence،في حين أن بعضها يعتمد على تجزئة الاختبار إلى مجموعتين متكافئتين من المفردات لتقدير الاتساق الداخلى لمفردات الاختبار Internal Consistency، والبعض الآخر يعتمد على تحليل البنية الداخلية للاختبار Internal Structure للتعرف على مدى تجانس مفردات الاختبار Tests Items Homogeneity مثل معامل كيودر وريتشاردسون Kuder - Richardson ومعامل ألفا لكرونباك Cronbach .

والشكل التخطيطى (٤ ـ ٤) التالى يوضح هذه الأنواع من معاملات الثبات:



شكل (٤ ـ ٤) يوضح الأنواع المختلفة من معاملات الثبات

(من تصميم المؤلف)

ويتضح من شكل (٤ ـ ٤) أن بعض أنواع معاملات الثبات يتطلب تطبيقين للاختبار أو صيغ مكافئة له، والبعض الآخر يتطلب تطبيقًا واحدًا.

وسوف نتناول هذه الأنواع المختلفة من معاملات الثبات فيما يلي بالمناقشة والتحليل ، ونبين كيفية تقدير قيم كل منها إمبيريقيا.

كيفية تقدير قيم أنواع معاملات الثبات :

سوف نتناول أولا كيفية تقدير قيم معاملات الثبات التي تعتمد على تطبيقين مختلفين لصيغ متكافئة من الاختبار أو للاختبار نفسه، ثم ننتقل إلى معاملات الثبات التي تعتمد على تطبيق الاختبار مرة واحدة.

معامل التكافؤ :

يعتمد تقدير الثبات اعتمادًا أساسيًا كما سبق أن أوضحنا على الارتباط بين مجموعتين من القياسات ، ولذلك فإن تقدير قيم معامل الثبات باستخدام صيغتين متكافئتين للاختبار يتفق منطقيًا مع هذا المفهوم.

غير أن المشكلة المتعلقة بذلك هي كيفية إعداد صيغ متكافئة أو متوازية لاختبار معين استنادًا إلى أساليب إجرائية محددة، فإحدى الصيغتين يمكن أن تختلف في محتواها وأنواع مفرداتها ودرجة صعوبتها عن الصيغة الأخرى مما يؤدي إلى اختلاف واضح في تباين الدرجات الحقيقية للفرد في كل منهما.

ولعل نموذج معاينة النطاق Domain Sampling Model الذي سبق أن أشرنا إليه يقدم إجابة نظرية لهذه المشكلة، حيث يفترض نطاقًا شاملاً للمفردات Items Domain تنتقى منه مفردات كل من صيغتي الاختبار انتقاءً عشوائيًا.

ونظرا لصعوبة إجراء ذلك عمليًا فإنه يمكن إعداد مواصفات تفصيلية دقيقة لنطاق المفردات بحيث تشمل حدود محتوى الاختبار، وتوزيعه على مفرداته، ونوع المفردات وعددها ودرجة صعوبتها، وأساليب انتقائها وفحصها ، وكذلك تعليمات الإجابه، وحدود زمنها، وأمثلة توضيحية، وجميع الجوانب الأخرى المتعلقة بتطبيق الاختبار وتصحيح الإجابات.

فإذا اتبعت هذه المواصفات في بناء صيغتي الاختبار مع مراعاة عدم تكرار المفردات في كل منهما، وكذلك عدم تداخل محتواها بدرجة كبيرة، وانتقاء مفردات كل منهما وفقا لهذه المواصفات، فإن ذلك يكون كافيًا لضمان تكافؤ الصيغتين، ويقال: إن الاختبارين في هذه الحالة متوازيان Parallel Tests. كما يمكن في مرحلة إعداد الاختبار تجميع عدد كبير من المفردات المتعلقة بالظاهرة المراد قياسها وتجريبها

ميدانيًا. فإذا كان محتوى الاختبار متجانسًا أو يشتمل على محتويات فرعية متجانسة، فإنه يمكن انتقاء المفردات بحيث يكون توزيع درجات صعوبتها في كل من صيغتي الاختبار له نفس الشكل، وكذلك توزيع قيم معامل الارتباط بين درجات المفردات والدرجة الكلية في كل منهما.

وربما يسهل إجراء أي من هاتين الطريقتين في اختبارات التحصيل التي تشتمل على مفردات تتطلب معرفة الكلمات أو الهجاء أو العمليات الحسابية، ولكن ربما يصعب إجراؤها في بعض اختبارات القدرات وكثير من مقاييس الشخصية ،حيث يكون إعداد صيغ متكافئة أو متوازية للاختبار أو المقياس أمرًا يتطلب جهدًا كبيرًا.

وعلى الرغم من هذه الصعوبات، إلا أن استخدام صيغ متكافئة أو متوازية للاختبار يكون ضروريًا في مواقف كثيرة. فعند تقويم المعلم لطلابه أو للمنهج الدراسي مثلاً فإنه يقوم بتطبيق اختبار في محتوى دراسي معين قبل بدء عملية التعليم، وعند انتهاء عملية التعليم يطبق اختبار آخر مكافئًا للاختبار الأول وذلك لتعرف النمو التحصيلي للطلاب.

فتوافر صيغتين للاختبار يتطلب معلومات تتعلق بمعامل تكافؤهما، وعلاوة على ذلك فإن تطبيق هاتين الصيغتين يجعل أثر الذاكرة والاستذكار من أجل الاختبار أقل مما لو اختبر الطلاب بالاختبار الأول نفسه قبل عملية التعليم وبعدها.

وكذلك في مواقف العلاج النفسي حيث يجرى تطبيق صيغتين متكافئتين لمقياس معين أحدهما قبل العلاج والآخر بعده، لتعرف فاعلية العلاج، فإن هذا يتطلب مؤشرات تتعلق بتكافؤ درجات صيغتي المقياس.

كذلك إذا كان اختبار معين يقيس أداء يتعلق بحل مشكلة محددة وتوصل الفرد إلى حلها في هذا الاختبار، فإن إعادة تطبيق الاختبار نفسه على الفرد سوف يجعله يتذكر خطوات الحل السابقة بطريقة آلية، وبذلك يكون الاختبار مقياسًا للتذكر وليس للقدرة على حل المشكلات.

أي أنه إذا توافرت صيغتان متكافئتان لاختبار أو مقياس معين (أ) ، (ب) فإن مشكلة الثبات في هذه الحالة تتعلق باتساق درجات كل من الصيغتين، ويصبح من الضروري تقدير قيمة معامل التكافؤ **Equivalent Coefficient** .

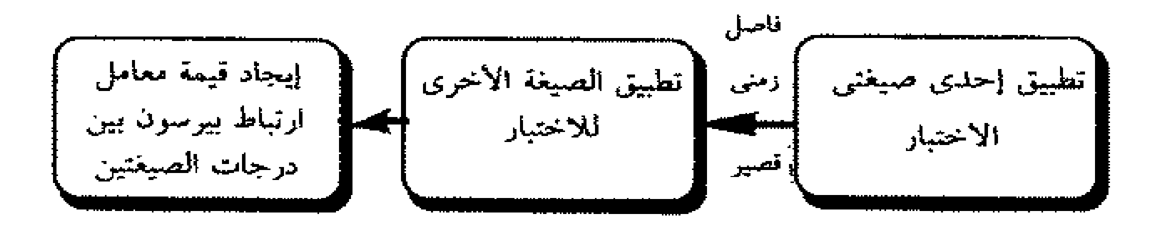
وهذا يتطلب تطبيق إحدى صيغتي الاختبار أو المقياس (الصيغة أ مثلا) في وقت ما ، وبعدها بقليل تطبق الصيغة الأخرى (الصيغة ب) ، أي أن الصيغتين تطبقان على مجموعة الأفراد واحدة تلو الأخرى أو بفاصل زمني قصير ، لكي لا تؤثر التغيرات التي ربما تحدث للأفراد المختبرين خلال هذا الفاصل الزمني في ثبات الدرجات باعتبار

أن هذه التغيرات غير متسقة، أى تعد من الأخطاء العشوائية. كما أن عوامل التذكر والمران والتعب لا تؤثر تأثيراً غير منتظم فى قيمة معامل التكافؤ.

وهذا الإجراء يشبه إلى حد ما وزن فرد مرتين متتاليتين بميزانين مختلفين فى اليوم نفسه، فهنا لا نتوقع أن يكون هناك اختلاف فى القيمتين يرجع إلى زيادة وزن الفرد. فإذا وُجد أى قدر من الاختلاف فإنما يرجع إلى اختلاف أداة القياس أى الميزان أو عملية القياس. ولإبطال التأثير الناجم عن ترتيب تطبيق كل من صيغتى الاختبار يمكن تقسيم مجموعة الأفراد إلى قسمين إحداهما تختبر بالصيغة (أ) تليها الصيغة (ب)، والأخرى تختبر بالصيغة (ب) تليها الصيغة (أ).

فالمصدر الأساسى للأخطاء العشوائية فى هذا النوع من الثبات يرجع إلى الفروق بين مفردات كل من صيغتى الاختبار أو المقياس التى ربما تكون نتيجة عدم تمثيل مفردات كل من الصيغتين للنطاق الشامل تمثيلاً كافيًا، أو عدم الالتزام بمواصفات هذا النطاق وبخاصة إذا كان هذا النطاق الشامل غير محدد تحديداً جيدًا، أو أن السمة التى يقيسها الاختبار سمة عريضة أو مركبة، مثل القدرة على التفكير الناقد أو الإبداع وكذلك ربما ترجع هذه الفروق إلى اختلاف إجراءات أو ظروف تطبيق كل من الصيغتين على الرغم من عدم وجود فاصل زمنى بين مرتى التطبيق.

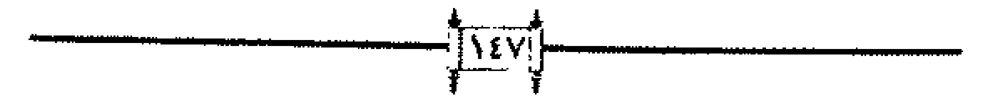
ويمكن تمثيل إجراءات هذه الطريقة بالشكل التخطيطى (٤ ـ ٥) التالى:



شكل (٤ ـ ٥) يوضح إجراءات تقدير قيمة معامل التكافؤ

وللحصول على قيمة تقديرية لمعامل التكافؤ نوجد معامل ارتباط بيرسون بين مجموعة درجات صيغتى الاختبار أو المقياس. وتعد هذه القيمة تقديرية؛ لأننا اعتمدنا على عينتين من المفردات ممثلتين لنطاقها الشامل كما سبق أن ذكرنا.

ولذلك فإن تفسيرها يتعلق بدقة قياس الاختبار للسمة المرجوة بغض النظر عن صيغة الاختبار المستخدمة أو ظروف تطبيقها. فكثير من القرارات التربوية والنفسية تعتمد على ذلك، فمثلاً إذا أراد مسئول عن العمل تقدير مهارة أحد العمال، فإنه لا



يقتصر على تقدير أدائه في موقف واحد أو بواسطة مشرف واحد، وإنما يعتمد على تقدير الأداء فى أكثر من موقف وأكثر من مشرف لكى يطمئن على اتساق التقديرات.

ومن الجدير بالذكر أن استخدام صيغتين متكافئتين يعد من أكثر الطرق دقة فى تقدير ثبات درجات الاختبارات والمقاييس التربوية والنفسية. غير أن كلفة بناء هاتين الصيغتين وما يتطلبه من جهد ووقت يقلل أحيانًا من استخدام هذه الطريقة ما لم تكن هناك صيغ متكافئة متوافرة. لذلك يفضّل كثير من المهتمين ببناء أدوات القياس استخدام طريقة تعتمد على إعادة تطبيق الاختبار نفسه لتقدير قيمة معامل الثبات الذى يُعرف بمعامل الاستقرار Stability Coefficient .

معامل الاستقرار:

يُستخدَم معامل الاستقرار في تقييم الخطأ الناجم عن تطبيق الاختبار مرتين متتاليتين، أى يعاد تطبيق الاختبار نفسه بحيث يكون هناك فاصل زمنى بين مرتى التطبيق. ولذلك يصلح هذا المعامل إذا كانت السمة التى يقيسها الاختبار لا تتغير كثيرا بمرور الزمن ، أى تكون مستقرة نسبيًّا.

فمثلاً لا يجوز استخدام هذا المعامل في تقدير ثبات درجات اختبار يقيس ذكاء الأطفال الصغار، أو درجات مقياس الحالة المزاجية، أو حالة القلق حيث إن هذه الخصائص تكون متذبذبة. أما إذا كانت السمة تتميز بالثبات النسبى أو الاستقرار، فإن اختلاف درجات الفرد فى مرتى تطبيق الاختبار، ربما يعزى إلى الأخطاء العشوائية للقياس. كما يصلح إذا أردنا اتخاذ قرارات لوضع خطط بعيدة الأمد مثل التنبؤ بالنجاح فى إحدى الكليات باستخدام اختبار الاستعداد الدراسى.

وقد أوضحنا أن معامل التكافؤ يتعلق بتقييم التأثير المحتمل لاختلاف مفردات كل من صيغتى الاختبار فى درجات كل منهما.

أما هذا المعامل فيتعلق بتقييم التأثير المحتمل لاختلاف الظروف المتعلقة بالفرد والبيئة المحيطة عند إعادة تطبيق الاختبار عليه.

ومن ميزات هذا النوع من الثبات أنه يسهل تقديره، وذلك بتطبيق الاختبار على مجموعة من الأفراد وإعادة تطبيقه عليهم بعد مرور مدة زمنية معينة فى ظروف مناسبة ومحددة، وإيجاد قيمة معامل ارتباط بيرسون بين مجموعتى الدرجات التى نحصل عليها فى مرتى التطبيق، وبذلك نحصل على قيمة تقديرية لمعامل الاستقرار Stability Coefficient،ويسمى هذا الإجراء طريقة إعادة الاختبار Test - Retest . والشكل التخطيطى (٤ ـ ٦) التالى يوضح ذلك :



شكل (٤ ـ ٦) يوضح إجراءات تقدير قيمة معامل الاستقرار

غير أن هناك بعض الاعتبارات التى ينبغى مراعاتها عند استخدام هذه الطريقة فى تقدير ثبات درجات الاختبارات والمقاييس. فمعامل الاستقرار لا يعكس الخطأ الناجم عن معاينة المحتوى فى الاختبار، كما أن هذا المعامل يتأثر تأثرًا ملحوظًا بالمدة الزمنية الفاصلة بين مرتى تطبيق الاختبار. فإذا كانت هذه المدة قصيرة جدًا فإن عوامل الذاكرة أو انتقال الأثر والتمرن ربما تؤدى إلى اتساق الدرجات فى مرتى التطبيق إلى حد كبير مما يؤدى إلى تضخيم قيمة معامل الاستقرار. فإعادة تطبيق الاختبار نفسه تؤدى فى كثير من الأحيان إلى تغيرات معينة فى الفرد ربما تسهم فى تحسين أدائه فيما يقيسه الاختبار كما هو الحال إذا طبق اختبار لقياس المهارة اليدوية لدى مجموعة من الأفراد، فإنهم يحصلون على درجات أفضل إذا ما أعيد تطبيق الاختبار نفسه عليهم نتيجة لاختبارهم فى المرة الأولى.

أما إذا كانت المدة الزمنية الفاصلة كبيرة بحيث تبلغ عدة أشهر مثلاً، فإننا نتوقع تدخل كثير من العوامل غير المرغوبة التى ربما تفسر اختلاف الدرجات فى مرتى التطبيق وبخاصة فى قياس التحصيل الدراسى، أو إذا كان المختبرون من الأطفال الصغار.

واختلاف هذه المدة الزمنية يؤدى إلى اختلاف قيم معامل الاستقرار حيث تعتمد هذه القيم على زيادة أو نقص هذه المدة. ويتضح ذلك إذا حاولنا تمثيل العلاقة بين قيمة معامل الاستقرار والمدة الزمنية الفاصلة بين مرتى التطبيق، فإننا سوف نلاحظ أن قيمة معامل الاستقرار تقل بزيادة هذه المدة. غير أن المدة الزمنية فى ذاتها ليست سببًا فى اختلاف الدرجات فى مرتى التطبيق، وإنما عامل المران أو الخبرة التى يكتسبها الفرد خلال هذه المدة.

لذلك ينبغى عند تقرير القيمة التقديرية لمعامل استقرار اختبار أو مقياس معين أن يذكر بالإضافة إلى هذه القيمة المدة الزمنية الفاصلة بين مرتى التطبيق، كما يجب أن تذكر بعض المعلومات أو البيانات من عينات مختلفة من الأفراد طبق عليهم الاختبار أو المقياس مرتين تفصل بينهم فترات زمنية مختلفة كلما أمكن ذلك.

وكذلك ينبغى عند الاطلاع على أدلة الاختبارات Test Manuals توجيه الاهتمام إلى المدة الزمنيـة الفاصلة التى فى ضوئها تـم تقدير قيمة معامـل الاتساق، وذلك قبل استخدام اختبار أو مقياس معين، وذلك للتأكد من أن درجات الاختبار سوف تكون مستقـرة أى ثابتة خلال المدة الزمنيـة التى يتطلبها مستخدم الاختبار فى دراسته. كما ينبغى تعرف الظروف أو المتغيرات التى ربـما تكون قد لعبت دوراً خلال المدة الفاصلة مثل تدرب المختبرين فى برامج معينة أو مشاهدة برامج التلفاز أو التحاقهـم ببرنامج تعليمى معين مما يؤثر تأثيراً ملحوظًا فى القيمة التقديرية لمعامل الاستقرار.

ولكن انخفاض قيمة هذا المعامل فى بعض الأحيان لا تعنى أن درجات الاختبار غير مستقرة، إذ ربما تـكون السمة المقـاسة قد تغيـرت على الرغم من افتراضنا أنها مستقرة نسبيًا.

لذلك يرى بعض علماء النفس المعاصرين مثل أتكينسون (Atkinson,1985) أن السمات الإنسانية الجوهرية مثل الدافعية تتذبذب بمرور الوقت. وعلى الرغم من أنه فى إطار طريقـة إعادة الاختـبار يعد هـذا التذبذب من الأخطـاء العشوائيـة، إلا أن هؤلاء العلماء يرون أنه يمكن التـنبؤ بالتغيرات التى تحدث فى دافعية الأفـراد باستخدام نماذج نظرية معينة على اعتبار أن هذه التغيرات ليست عشوائية، وإنما تغيرات منتظمة.

وعمومًا فإن قلـة مصادر الأخطاء العشوائية فى هذه الطـريقة يمكن أن تؤدى إلى قيم تقديرية مرتفعة نسبيًا لمعامل الاستقرار.

ولذلك فإنه يمكن الإفادة من ميزات كل من معامل الاستقرار ومعامل التكافؤ فى تقدير ثبـات درجات الاختبارات، ويسـمى المعامل عندئـذ معامل الاستقـرار والتكافؤ Stability and Equivalence.

معامل الاستقرار والتكافؤ:

ذكرنا فيما سبق أن معامل التكافؤ يتطلب تطبيق صيغتى الاختبار فى الوقت نفسه تقريبًا أو بفـاصل زمنى قصير مـن أجـل اتخاذ قرارات أو التوصل إلى تنبـؤات قصيرة الأمد، مثل تقديرات المشرفين على العمل أثنـاء أداء العاملين لمهامهم. أو تتبع التقدم الدراسى للطلاب خلال مدة دراسية محدودة.

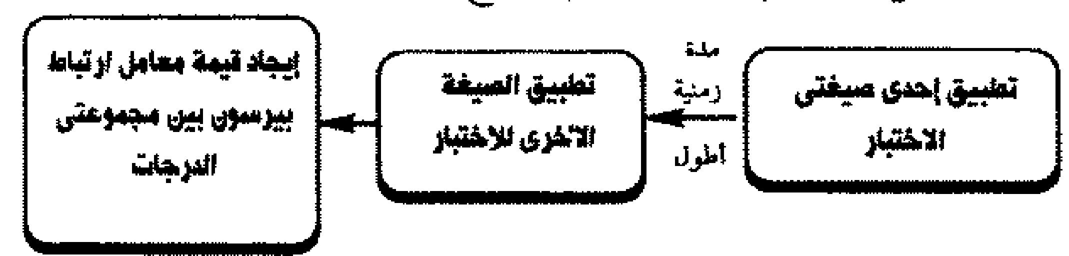
ولكننا أحيانًا نود استـخدام نتائج الاختبارات والمقاييـس فى التوصل إلى تنبؤات بعيدة الأمد، مثل التنبؤ بنجاح الطلاب فى الكليات باستخدام درجات اختبارات القبول، أو درجات الاستعداد الدراسى أو المعدل التراكمى فى الثانوية العامة، حيث يفترض أن هذه الدرجات مستقرة، أى أننا نهتم هنا بمعامل استقرار الاختبار.

ولكننا فى أحيان أخرى ربما لا يقتصر اهتمامنا على التنبؤات بعيدة الأمد فقط، وإنما أيضًا بالاستدلال على نطاق شامل لمحتوى مجال مهنى أو دراسى معين. فمثلاً إذا اعتمدنا فى قياس كثير من المفاهيم السيكلوجية مثل الذكاء، والميول، والاتجاهات على مجموعة محددة من المفردات الاختبارية، فإن هذه المفاهيم تصبح قليلة الجدوى.

ولكننا نود التحقق من إمكانية قياس هذه المفاهيم بمجموعات مختلفة من المفردات المستمدة من نطاق شامل Items Domain بحيث تكون درجات الأفراد الذين يُختبرون بكل مجموعة منها متطابقة أو متقاربة. لذلك يفضل فى هذه الحالة استخدام معامل الاستقرار والتكافؤ Stability and Equivalence .

ويمكن تقدير قيمة هذا المعامل بطريقة مماثلة لطريقة تقدير قيمة معامل التكافؤ، غير أنه يجب أن تكون هناك مدة زمنية أكبر تفصل بين مرتى التطبيق، ونوجد معامل ارتباط بيرسون بين مجموعة الدرجات التى نحصل عليها.

والشكل التخطيطى (٤ ـ ٧) التالى يوضح ذلك:



شكل (٤ ـ ٧) يوضح إجراءات تقدير قيمة معامل الاستقرار والتكافؤ

ونظرًا لأن هذه الطريقة تجمع بين الطريقتين السابقتين فإنها تؤدى إلى قيم تقديرية لمعامل الثبات أقل من نظيرتها فى أى من الطريقتين السابقتين؛ وذلك لأنها تجمع بين الأخطاء العشوائية التى تؤثر فى كل منهما.

فمعامل الاستقرار والتكافؤ يعكس الأخطاء العشوائية الناجمة عن اختلاف مفردات صيغتى الاختبار، وكذلك اختلاف الدرجات نتيجة التغيرات التى تحدث للأفراد المختبرين أو تذبذب السمة التى يقيسها الاختبار. ولذلك يعد هذا المعامل أكثر المعاملات تدقيقًا، وقيمه تعد بمثابة الحد الأدنى لتقدير معامل الثبات.

وسوف نقدم فيما يلى مثالاً يوضح كيفية إيجاد القيمة التقديرية لكل من معامل التكافؤ ، ومعامل الاستقرار، ومعامل التكافؤ والاستقرار:

نفترض أننا طبقنا صيغتين متكافئتين (أ) ، (ب) من اختبار نفسى على مجموعة تتكون من ٢٠ فردًا فى أوقات مختلفة، فالصورة (أ) طبقت فى يوم ما تلتها الصورة

(ب) في اليوم نفسه، أما الصورة (أ) فقد أعيد تطبيقها بعد انقضاء أسبوعين، وسوف نرمز لهذه التطبيقات الثلاثة بالرموز: $أ_١$ ، $ب_١$ ، $أ_٢$ على الترتيب ، وفيما يلي الدرجات المناظرة لها :

$أ_١$ ١٦ ١٦ ١٨ ٢٠ ١٠ ١٤ ٢٤ ١٢ ٣٠ ٢٨ ٢٤ ٢٢ ٢٠ ٢٠
١٨ ١٨ ٢٦ ٢٠ ٢٢ ٢٠

$ب_١$ ١٢ ١٦ ١٨ ٢٠ ٢٠ ١٨ ٢٠ ١٢ ١٤ ١٨ ٢٦ ٢٨ ١٦ ٣٢
١٤ ٢٠ ٢٢ ٢٨ ٢٤ ٢٢

$أ_٢$ ٢٤ ٢٢ ١٦ ١٤ ٢٠ ٢٠ ٢٢ ٣٠ ٢٢ ٢٤ ١٦ ١٨ ٢٢ ٣٠
٢٨ ٢٤ ٢٠ ٢٦ ٣٢ ٢٦

فلإيجاد القيمة التقديرية لمعامل التكافؤ نستخدم نتائج الاختبارين ($أ_١$ ، $ب_٢$) ونوجد قيمة معامل ارتباط بيرسون بين درجاتهما باستخدام الصيغة (٣ ـ ١٠)، وهذه تتطلب إيجاد قيم مجـ س ، مجـ $س^٢$ ، مجـ س ص لكل من الاختبارين كالتالي:

الاختبار	$أ_١$	$ب_١$
مجـ س	٣٩٨	٤٠٠
مجـ $س^٢$	٨٤٠٤	٨٥٧٦
مجـ س ص	٨٤٧٢	

$$ر = \frac{٤٠٠ \times ٣٩٨ - ٨٤٧٢ \times ٢٠}{\sqrt{[\,^٢(٤٠٠) - ٨٥٧٦ \times ٢٠\,][\,^٢(٣٩٨) - ٨٤٠٤ \times ٢٠\,]}}$$

$$= ٠,٩٦٩$$

ولإيجاد القيمة التقديرية لمعامل الاستقرار نستخدم نتائج الاختبارين $أ_١$ ، $أ_٢$ ونكرر ما سبق كالتالي:

الاختبار	أ ١	أ ٢
مجـ س	٣٩٨	٤٥٦
مجـ س ٢	٨٤٠٤	١٠٨٥٦
مجـ س ص	٩٥٠٠	

$$ر = \frac{٢٠ \times ٩٥٠٠ - ٣٩٨ \times ٤٥٦}{\sqrt{[٢٠ \times ٨٤٠٤ - (٣٩٨)^٢][٢٠ \times ١٠٨٥٦ - (٤٥٦)^٢]}}$$

= ٠,٩٠٣

ولإيجاد القيمة التقديرية لمعامل الاستقرار والتكافؤ نستخدم نتائج الاختبارين $ب_١$، $أ_٢$ ونكرر الخطوات نفسها كالتالى :

الاختبار	ب ١	أ ٢
مجـ س	٤٠٠	٤٥٦
مجـ س ٢	٨٥٧٦	١٠٨٥٦
مجـ س ص	٩٥٨٨	

$$ر = \frac{٢٠ \times ٩٥٨٨ - ٤٠٠ \times ٤٥٦}{\sqrt{[٢٠ \times ٨٥٧٦ - (٤٠٠)^٢][٢٠ \times ١٠٨٥٦ - (٤٥٦)^٢]}}$$

= ٠,٩٠٩

من هذا يتضح أن ترتيب القيم التقديرية لمعامل الثبات هو : ٠,٩٦٩ ، ٠,٩٠٩ ، ٠,٩٠٣ حيث يلاحظ أن قيمة معامل التكافؤ ٠,٩٦٩ تعد أعلى القيم الثلاث نظرًا لتكافؤ درجات الاختبارين ($أ_١$) ، ($ب_١$) ، فمتوسط درجات كل منهما: ١٩,٩ ، ٢٠,٠ ، والانحراف المعيارى للدرجات ٢,٤٦ ، ٢,٦٨ على الترتيب.

أى أن هناك تقاربًا كبيرًا بين متوسطى درجات كل منهما وانحرافهما المعيارى وبذلك قلت الأخطاء العشوائية التى تعزى إلى اختلاف عينة مفردات كل من الاختبارين، حيث إن نسبة التباين الحقيقى (٠,٩٦٩)٢ ، أى ٩٤٪ تقريبًا مما يدل على توازى الاختبارين.

كما يلاحظ أن قيمة كل من معامل الاستقرار، ومعامل الاستقرار والتكافؤ ٠,٩٠٩ ، ٠,٩٠٣ على الترتيب ، وهما قيمتان متقاربتان مما يدل على اتساق درجات الاختبار (أ) فى مرتى التطبيق، وكذلك تكافؤ درجات الاختبار (أ) والاختبار (ب) مما يدل على توازيهما كما سبق أن أشرنا.

وهذا أيضًا يدل على أن السمة التى تقيسها كل من صيغتى الاختبار (أ) مستقرة نسبيًا خلال المدة الزمنية الفاصلة على الأقل. ونظرًا لاختلاف متوسطى كل من الاختبارين $أ_١$ ، $أ_٢$ حيث إن هذين المتوسطين ١٩,٩ ، ٢٢,٨ على الترتيب ، فإن هذا يدل على أن الأخطاء العشوائية الناجمة عن تأثير المران والتدرب والتى ظهرت فى تحسن الدرجة فى الاختبار $أ_٢$ (أى بعد إعادة تطبيقه) قد انعكست فى معامل الاستقرار حيث بلغت قيمته التقديرية ٠,٩٠٣ ، ومع هذا فإن نسبة التباين الحقيقى فى الدرجات (٠,٩٠٣)٢ ، أى ٨٢٪ تقريبًا وهى نسبة مرتفعة،حيث إن نسبة تباين الخطأ فى الدرجات تبلغ ١٨٪ فقط.

معامل الاتساق الداخلى:

تتطلب الطرق الثلاث السابقة تطبيق اختبار أو صيغ متكافئة منه مرتين بفواصل زمنية متباينة، غير أنه فى بعض الأحيان يصعب بناء صيغتين متكافئتين أو الحصول عليهما، أو قد يصعب تطبيق الاختبار مرتين، فعندئذ يفضل تقدير الثبات باستخدام اختبار واحد وتطبيقه مرة واحدة.

فالمعلم لا يمكنه إعادة تطبيق الاختبار التحصيلى نفسه مرتين متتاليتين لتقويم طلابه، كما لا يمكنه بناء صيغ متعددة من الاختبار، ولذلك يمكنه تقدير ثبات الاختبار باستخدام معامل الاتساق الداخلى Internal Consistency .

وهذه الطريقة تماثل طريقة الصيغتين المتكافئتين، غير أنها لا تهدف لتقييم تكافؤ الصيغتين نظرًا لأنها تعتمد على تطبيق اختبار واحد ثم تجزئته إلى نصفين متكافئين،

وإيجاد معامل ارتباط بيرسون بين درجات كل من النصفين بعد الانتهاء من تطبيق الاختبار، ولذلك فهى تهتم بتقييم الاتساق الداخلى لمفردات الاختبار.

ويتم تقدير درجات كل من نصفى الاختبار كما لو كان كل منهما اختباراً منفصلاً. غير أن المشكلة الأساسية تتعلق بكيفية تجزئة الاختبار إلى نصفين متكافئين أو متوازيين وبخاصة إذا كان الاختبار يشتمل على مفردات غير متجانسة فى محتواها أو يقيس سمة مركبة أو عدة سمات، مثل بطاريات اختبارات الاستعداد العقلى أو المهارات الأساسية مثلاً، وكذلك إذا كان الاختبار موقوتًا، أو تتباين مفرداته فى درجة صعوبتها أو متزايدة الصعوبة، أو متداخلة فيما بينها.

لذلك فإن هذه الطريقة تتطلب الدقة والحرص عند تجزئة الاختبار إلى نصفين متكافئين ، حيث إنه يمكن إجراء هذا التقسيم بطرق متعددة.

ومع هذا فإنه إذا كان الاختبار غير متجانس فى محتواه، فإنه يمكن تجميع المفردات المتجانسة معًا أو تقسيمه إلى اختبارات فرعية متجانسة فى محتواها وتجزئة كل اختبار فرعى إلى نصفين متكافئين، وضم أجزاء كل نصف على حدة.

أما إذا كانت مفردات الاختبار متزايدة الصعوبة فإنه يمكن تجزئة الاختبار إلى نصفين متكافئين يشتمل أحدهما على المفردات الفردية والآخر على المفردات الزوجية.

ولكن من الضرورى فى جميع الحالات التأكد من تشابه مضمون المفردات التى تكون النصفين وفق خطة منظمة، وكذلك التحقق من تساوى متوسط درجات كل من النصفين والانحراف المعيارى.

كذلك لا يجوز استخدام هذه الطريقة إذا كان الاختبار من نوع اختبارات السرعة Speed Tests حيث يتطلب أن يجيب الفرد عن عدد كبير من المفردات السهلة نسبيًا فى زمن محدد، مثل اختبار القدرة العددية الذى يشتمل على عدد كبير من المفردات التى تتطلب عمليات جمع بسيط فى زمن وجيز. واختلاف الأفراد فى معدل سرعة إجابتهم يجعل تجزئة الاختبار إلى نصفين متكافئين غير مناسب بل وغير مطلوب. لذلك يفضل استخدام معامل الاتساق الداخلى فى تقدير ثبات اختبارات القوة Power Tests.

ولكن إذا كان من الضرورى تقدير ثبات درجات اختبار موقوت،أى يعتمد على السرعة فإنه يمكن إجراء ذلك بإيجاد معامل الارتباط بين اختبارين لكل منهما زمنه المحدد ويقيسان السمة نفسها بدلاً من تجزئة الاختبار إلى نصفين متكافئين.

ويقترح ستانلى Stanley فيما يتعلق باختبارات القدرات الموقوتة التى لا تهدف أساسًا إلى جعل سرعة الأفراد المستجيبين تؤثر فى درجاتهم تأثيراً كبيراً ترتيب مفردات

الاختبار من حيث درجة صعوبتها أى من السهل إلى الصعب، والسماح بوقت كاف يمكّن جميع الأفراد من محاولة إجابة كل مفردة قدر استطاعته بالترتيب التى ترد به فى الاختبار إلى نصفين متكافئين بالنسبة لزمن الإجابة وليس بالنسبة للمفردات، أى إعطاء زمن محدد لكل من نصفى الاختبار.

ومن المشكلات الأخرى لهذه الطريقة أن الدرجات تعتمد على نصف عدد مفردات الاختبار. ونظراً لأن القيمة التقديرية لمعامل الثبات تتأثر بطول الاختبار أى عدد مفرداته، فإن قيمة معامل الاتساق الداخلى الناتجة عن إيجاد معامل ارتباط بيرسون بين نصفى الاختبار تكون أقل مما لو استخدم العدد الكلى لمفردات الاختبار، أى أنها تدل على ثبات درجات نصف الاختبار.

ولتقدير ثبات الاختبار ككل ينبغى افتراض إطالة الاختبار إلى الضعف، وإجراء تعديل على قيمة معامل ارتباط بيرسون التى حصلنا عليها فى ضوء هذه الإطالة.

وقد اشتق كل من سبيرمان وبراون **Spearman - Brown** صيغة رياضية عامة لتقدير معامل ثبات الاختبار إذا زاد طوله أو نقص (ك) من المرات، وذلك فى إطار مبادئ النظرية الكلاسيكية للاختبارات، وهذه الصيغة هى :

$$\text{معامل الثبات التقديرى } ر_{ك ك} = \frac{ك \times ر_{س س}}{1 + (ك - 1)\, ر_{س س}} \quad \ldots\ldots\ldots (4 - 7)$$

حيث $ر_{س س}$ ترمز إلى معامل ثبات درجات الاختبار الأصلى قبل الإطالة.

، $ر_{ك ك}$ ترمز إلى معامل الثبات التقديرى لدرجات الاختبار بعد إطالته (ك) من المرات.

وفى حالة التجزئة النصفية، فإن ك = ٢، وتصبح صيغة سبيرمان وبراون كالتالى:

$$ر_{11} = \frac{2\, ر_{21}}{1 + ر_{21}} \quad \ldots\ldots\ldots\ldots (4 - 8)$$

حيث $ر_{11}$ ترمز إلى تقدير معامل ثبات درجات الاختبار ككل.

، $ر_{21}$ ترمز إلى قيمة معامل ثبات درجات نصف الاختبار.

وتفترض هذه الصيغة تساوى تباين درجات كل من نصفى الاختبار.

فإذا كان معامل الارتباط بين نصفى الاختبار = ٠,٨٠ ، فإن قيمة معامل ثبات درجات نصف الاختبار = ٠,٨٠ ، وتصبح القيمة التقديرية لمعامل ثبات درجات الاختبار ككل أى إذا زاد طوله إلى النصف كالتالى:

$$ر_{١١} = \frac{٢ \times ٠,٨٠}{١ + ٠,٨٠} = \frac{١,٦٠}{١,٨٠} = ٠,٨٩$$

ويلاحظ أن هذه القيمة التقديرية لمعامل ثبات درجات الاختبار ككل قد زادت عن القيمة المناظرة لمعامل ثبات درجات نصف الاختبار.

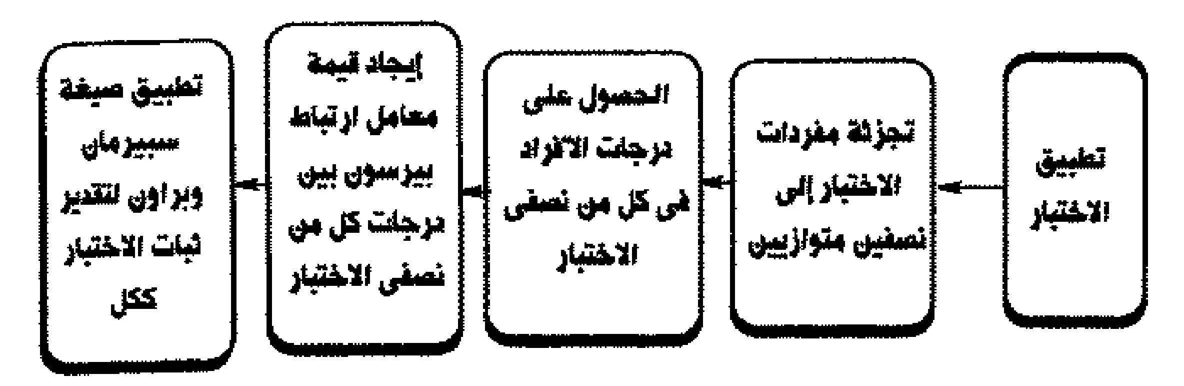
والجدول التالى (٤ ـ ١) يوضح تأثير استخدام تصحيح سبيرمان وبراون فى القيمة التقديرية لمعامل الاتساق الداخلى لقيم تتراوح بين ٠,١٠ ، ٠,٩٠ قبل إجراء التصحيح:

قيم المعامل قبل التصحيح	٠,١٠	٠,٢٠	٠,٣٠	٠,٤٠	٠,٥٠	٠,٦٠	٠,٧٠	٠,٨٠	٠,٩٠
قيم المعامل بعد التصحيح	٠,١٨	٠,٣٣	٠,٤٦	٠,٥٧	٠,٦٧	٠,٧٥	٠,٨٢	٠,٨٩	٠,٩٥
مقدار التغير فى قيم المعامل	٠,٠٨	٠,١٢	٠,١٦	٠,١٧	٠,١٧	٠,١٥	٠,١٢	٠,٠٩	٠,٠٥

جدول (٤ ـ ١) يوضح القيم التقديرية لمعامل الاتساق الداخلى
قبل وبعد التصحيح باستخدام صيغة سبيرمان وبراون

ويتضح من جدول (٤ ـ ١) أن مقدار التغير فى قيم معامل الاتساق الداخلى تزداد بعد إجراء تصحيح سبيرمان وبراون، غير أن الزيادة تكون أكبر لقيم المعامل المحصورة بين ٠,٣٠ ، ٠,٦٠ أى القيم الوسطى.

ويمكن توضيح إجراءات تقدير قيمة معامل الاتساق الداخلى لاختبار أو مقياس معين باستخدام طريقة التجزئة النصفية Split-Half Method بالشكل التخطيطى (٤ـ٨) التالى :



شكل (٤ ـ ٨) يوضح إجراءات تقدير قيمة معامل الاتساق الداخلى للاختبار

وقد توصل فلانجان Flanagan إلي صيغة أخرى تختلف عن صيغة سبيرمان وبراون ولا تتطلب توافر شرط تساوى تباين درجات كل من نصفى الاختبار ، وهذه الصيغة هى :

$$ر_{س س} = ٢ [١ - \frac{ع^٢_{أ} + ع^٢_{ب}}{ع^٢_{س}}] \qquad (٤ ـ ٩)$$

حيث ($ر_{س س}$) ترمز إلى معامل ثبات الاختبار ككل.

، ($ع^٢_{أ}$) ترمز إلى تباين درجات أحد نصفى الاختبار.

، ($ع^٢_{ب}$) ترمز إلى تباين درجات النصف الآخر.

، ($ع^٢_{س}$) ترمز إلى تباين الدرجات الكلية فى الاختبار.

وتطبيق هذه الصيغة يتطلب إيجاد تباين درجات كل من نصفى الاختبار، وكذلك تباين درجاته الكلية باستخدام الصيغة (٣ ـ ٥) التى سبق أن أوضحناها فى الفصل الثالث.

والقيم التى نحصل عليها باستخدام هذه الصيغة تقترب من القيم الناتجة عن صيغة سبيرمان وبراون وبخاصة إذا تساوى تباين كل من نصفى الاختبار أى :

$$ع^٢_{أ} = ع^٢_{ب}$$

ولتوضيح كيفية استخدام هذه الصيغة فى إيجاد القيمة التقديرية لدرجات اختبار نفترض أننا قمنا بتجزئته إلى نصفين أحدهما يشتمل على المفردات الفردية والآخر يشتمل على المفردات الزوجية. ثم حسبنا قيمة تباين درجات كل من نصفى الاختبار

(التباين يساوى مربع الانحراف المعيارى) ودرجات الاختبار ككل فوجدنا أنها تساوى ٢,١٦ ، ١,٨٥ ، ٧,١٨ على الترتيب. وبالتعويض عن هذه القيم فى صيغة فلاناجان (٤ ـ ٩) نجد أن :

ر~سس~ = ٢ [١ − (٢,١٦ + ١,٨٥) / ٧,١٨]

= ٢ [١ − ٤,٠١ / ٧,١٨]

= ٠,٨٨٢

ويمكن أن نحصل على هذه القيمة نفسها بإيجاد معامل الارتباط بين درجات نصفى الاختبار وتصحيح القيمة الناتجة باستخدام صيغة سبيرمان وبراون كما سبق.

ومن مزايا الصيغة العامة لسبيرمان وبراون أنه يمكن باستخدامها تقدير طول الاختبار أى عدد مفرداته اللازم لجعل ثبات درجاته قيمة معينة إذا علمنا قيمة معامل ثبات درجات الاختبار الأصلى.

فمثلاً إذا اشتمل اختبار على ١٥ مفردة، وقيمة معامل ثبات درجاته = ٠,٧٠ فإنه يمكن تقدير طول الاختبار وعدد مفرداته التى تجعل قيمة الثبات ٠,٨٠ وذلك بتطبيق الصيغة العامة (٤ـ٧) والتعويض فيها عن قيمة ر~كك~ = ٠,٨٠ ، ر~سس~ = ٠,٧٠ ، وحل المعادلة الناتجة لإيجاد قيمة (ك) كالتالى:

٠,٨٠ = (ك × ٠,٥٠) / (١ + (ك ـ ١) (٠,٥٠))

٠,٨٠ = ٠,٥٠ ك / (٠,٥٠ ك + ٠,٥٠)

٠,٤٠ ك + ٠,٤٠ = ٠,٥٠ ك

٠,١٠ ك = ٠,٤٠

ك = ٤ .

لهذا فإنه ينبغي إطالة الاختبار إلى أربعة أمثال طوله الأصلي ، أي أن عدد المفردات يجب أن يزداد بحيث يصبح ٤ × ١٥ = ٦٠ مفردة.

ولكن ينبغي ملاحظة أن هذه الصيغة العامة تتطلب أن تكون المفردات التي ستضاف إلى الاختبار الأصلي مستمدة من نطاق المفردات الذي تمثله مفردات هذا الاختبار ، كما ينبغي مراعاة الواقعية في اختيار القيمة المتوقعة (ك) لإطالة الاختبار من حيث الزمن اللازم للإجابة، وطبيعة عينة الأفراد المختبرين.

وتفسير القيم التقديرية لمعامل الاتساق الداخلي يماثل تفسير قيم معامل التكافؤ، حيث إن كل من نصفي الاختبار يطبقان في آن واحد كما هو الحال في الصيغتين المتكافئتين للاختبار.

والأخطاء العشوائية التي تدخل في تقدير معامل الاتساق الداخلي تتعلق جزئيًا بأخطاء معاينة محتوى كل من نصفي الاختبار، ولكن الخطأ الناجم عن عدم اتساق الدرجات عبر مدة زمنية فاصلة ليس له أي تأثير نظرًا لأن نصفي الاختبار يطبقان معًا.

ولذلك فإن هذه الطريقة تؤدي عادة إلى قيم تقديرية لمعامل الثبات أكبر من حقيقتها.

معامل التجانس لكيودر وريتشاردسون :

ذكرنا فيما سبق أن مشكلة تقدير قيمة معامل الاتساق الداخلي للاختبار بطريقة التجزئة النصفية تتعلق بكيفية تجزئة الاختبار إلى نصفين متكافئين، فاختلاف طريقة التجزئة يؤدي إلى اختلاف في القيمة التقديرية لمعامل الثبات ، كما أن هذه الطريقة أيضًا تتطلب تحقق شرط تساوي تباين درجات كل من نصفي الاختبار. غير أن هناك طرقًا أخرى هدفت لمعالجة أوجه قصور طريقة التجزئة النصفية، ومن بين هذه الطرق شائعة الاستخدام طريقة كيودر وريتشاردسون Kuder - Richardson ، وطريقة كرونباك Cronbach .

وتهدف طريقة كيودر وريتشاردسون Kuder - Richardson للتوصل إلى قيمة تقديرية لمعامل ثبات الاختبارات غير الموقوتة أي اختبارات القوة Power Tests،والتي تكون درجات مفرداتها ثنائية أي إما واحد صحيح أو صفر، مثل مفردات الاختيار من متعدد، أو مفردات الصواب والخطأ.

وتتغلب هذه الطريقة على مشكلة تعدد طرق التجزئة النصفية للإختبار بأن توجد قيمة تقديريه لمتوسط قيم معامل الارتباط بين كل من نصفي الاختبار لجميع طرق التجزئة الممكنة دون أن نقوم بهذه التجزئة فعلاً فعلى الرغم من أن جميع الطرق السابقة استخدمت الأساليب الإحصائية في تقدير قيمة معامل الثبات إلا أن طريقة كيودر وريتشاردسون تعد اسلوبا إحصائيًا في ذاتها، فتقدير متوسط قيم معامل



الارتباط بين كل من نصفى الاختبار اعتمد على المبادئ الإحصائية للنظرية الكلاسيكية للاختبارات.

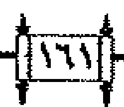
وتؤكد هذه الطريقة العلاقات القائمة بين المفردات التى يشتمل عليها الاختبار أكثر من تأكيدها اتساق درجات الاختبار بمرور الزمن أو تكافؤها على الرغم من اختلاف صيغ الاختبار.

وتعد طريقة التجزئة النصفية التى سبق ذكرها حالة خاصة من هذه الطريقة. فالاتساق الداخلى لمفردات الاختبار يعنى درجة الارتباط الموجب بين هذه المفردات. ولذلك يفضل كيودر وريتشاردسون تسمية ذلك تجانس المفردات Item Homogeneity، أى اتساق الأداء عبر جميع مفردات الاختبار.

ولذلك يرى براون Brown أن تجانس الاختبار يمكننا من التنبؤ بأداء الفرد فى مفردات الاختبار إذا علمنا أداءه فى إحدى هذه المفردات، وهذا يتطلب أن يقيس الاختبار سمة أحادية البعد Unidimensional ، أى تشتمل على متغير واحد.

وتجانس مفردات الاختبار يعد شرطًا أساسيًا للاختبار الذى يقيس سمة أحادية البعد. ولذلك يُنصح القائم ببناء أداة قياس أن يراعى عند استخدام طريقة كيودر وريتشاردسون هذا الشرط إلى جانب الشروط الأخرى التى ذكرناها. فإذا اشتملت أداة القياس على اختبارات فرعية غير متجانسة فإنه يمكن تقدير ثبات درجات كل منها على حدة. ولذلك فإن معامل الثبات الذى نحصل عليه باستخدام هذه الطريقة يسمى معامل التجانس Homogeneity Coefficient .

ونظرًا لتعدد الحالات المتعلقة بالخصائص الإحصائية لمفردات الاختبارات التى اهتم كيودر وريتشاردسون بدراستها، فقد تعددت الصيغ التى توصلا إليها. غير أن أكثر هذه الصيغ شيوعًا واستخدامًا فى تقدير قيمة معامل تجانس درجات مفردات الاختبارات الصيغة ٢٠ Kuder - Richardson Formula 20 (KR - 20) ، والصيغة ٢١ Kuder - Richardson Formula 21 (KR - 21) ، ولذلك سوف نقتصر على توضيح كل من هاتين الصيغتين.



الصيغة (٢٠) لكيودر وريتشاردسون :

تستخدم الصيغة (٢٠) إذا كانت درجات المفردات ثنائية (صفر ، ١) فقط، وتوافرت الشروط سالفة الذكر، وهذه الصيغة كما يلي :

$$\text{معامل التجانس} = \frac{ن}{ن - ١} \times \left(\frac{ع^٢ - \text{مجـ س ص}}{ع^٢} \right) \quad .. \ (٤-١٠)$$

حيث (ن) ترمز إلى عدد مفردات الاختبار.

، ($ع^٢$) ترمز إلى تباين الدرجات الكلية في الاختبار (مربع الانحراف المعياري).

، (س) ترمز إلى نسبة عدد الأفراد الذين أجابوا عن أي مفردة إجابة صحيح (درجة صعوبة المفردة).

، (ص) ترمز إلى نسبة عدد الأفراد الذين أجابوا عن أي مفردة إجابة خطأ.

أي أن ص = ١ − س ، (مجـ س ص) ترمز إلى مجموع تباين درجات مفردات الاختبار.

وبالطبع لن نناقش البرهان الرياضي لهذه الصيغة حيث إنه يتطلب مراجعة لمبادئ وفرضيات النظرية الكلاسيكية للاختبارات ، ولكن ما يهم القائمين ببناء الاختبارات والمقاييس ومستخدميها هو كيفية استخدام هذه الصيغة استخداما صحيحًا مع تحقق الشروط التي تتطلبها.

وإذا نظرنا إلى الصيغة (٤ - ١٠) نجد أن كسر النسبة الثانية

$$\left(\frac{ع^٢ - \text{مجـ س ص}}{ع^٢} \right) \text{ يمكن كتابته } \left(١ - \frac{\text{مجـ س ص}}{ع^٢} \right)$$

فإذا كانت مجـ س ص = $ع^٢$ ، أي أن مجموع تباين درجات مفردات الاختبار يساوي تباين الدرجات الكلية في الاختبار ، فإن هذا الكسر يصبح مساويًا صفرًا ، وبذلك تنعدم القيمة التقديرية لمعامل التجانس .

ولذلك إذا أردنا أن تكون قيمة معامل ثبات الاختبار مرتفعة ينبغي أن تكون (مجـ س ص) أكبر من ($ع^٢$)، أي ينبغي أن يكون تباين الدرجات الكلية في الاختبار أكبر من مجموع تباين درجات مفردات الاختبار. وهذا لا يتأتى إلا إذا كان هناك ارتباط مرتفع بين المفردات، أي تقيس جميعها سمة واحدة، وهو أحد الشروط الأساسية لهذا المعامل.



ولتوضيح ذلك نفترض أن لدينا مقياسًا يشتمل على ٢٠ عبارة تتعلق بسمة الخجل ، وتتطلب هذه العبارة من الفرد المستجيب أن يذكر ما إذا كانت العبارة تنطبق عليه (يحصل على الدرجة واحد صحيح) ، أم لا تنطبق (يحصل على الدرجة صفر) وطبق المقياس على ٤٠ فردًا.

فلإيجاد القيمة التقديرية لمعامل التجانس لهذا المقياس باستخدام الصيغة (٢٠) لكيودر وريتشاردسون ، فإنه ينبغي أولا التحقق من توافر شروط هذه الصيغة.

ونظرًا لأن جميع العبارات تقيس سمة يفترض أنها أحادية البعد، والمقياس لا يعتمد على السرعة ، ومفرداته ثنائية الدرجة، فإننا نستطيع استخدام هذه الصيغة وفقًا للخطوات التالية :

الخطوة الأولى : نكّون مصفوفة درجات المفردات لمجموعة الأفراد، حيث تحتوى خلايا هذه المصفوفة على الدرجات صفر ، ١. فإذا وجد الفرد الأول مثلا أن العبارة الأولى لا تنطبق عليه ، فإنه يحصل على الدرجة صفر، وتدون فى الخلية المقابلة للفرد الأول والمفردة الأولى فى المصفوفة، وهكذا فى بقية المفردات والأفراد.

الخطوة الثانية : نجمع الدرجات الكلية لكل فرد على حدة ، وكذلك المجموع الكلى لدرجات الأفراد، ونحسب قيمة كل من المتوسط والانحراف المعيارى لهذه الدرجات الكلية، فيكون تباين هذه الدرجات (ع^{٢}).

الخطوة الثالثة : نوجد قيمة (س) وذلك بقسمة عدد الأفراد الذين حصلوا على الدرجة واحد صحيح فى كل عبارة من عبارات المقياس على العدد الكلى لأفراد المجموعة.

الخطوة الرابعة : نوجد قيمة (ص) وذلك بطرح قيمة (س) التى حصلنا عليها فى الخطوة الثالثة من الواحد الصحيح.

الخطوة الخامسة : نوجد حاصل ضرب قيم (س) ، (ص) المستقابلة، ونجمع نواتج حاصل الضرب لجميع العبارات لنحصل على مجـ (س ص).

الخطوة السادسة : نعوض عن قيمة كل من (ع^{٢})(مجـ س ص)،(ن)فى الصيغة (٤ ـ ١٠) السابقة للحصول على القيمة التقديرية لمعامل التجانس.

فمثلاً إذا وجدنا أن ع^{٢} = ٣,٩ ، مجـ س ص = ١,٤

$$\text{فإن معامل التجانس} = \frac{٢٠}{٢٠ - ١} \times \frac{٣,٩ - ١,٤}{٣,٩}$$

$$= \frac{٢٠}{١٩} \times \frac{٢,٥}{٣,٩} = ٠,٦٨$$

وتفسير هذه القيمة لا يختلف عن تفسير القيم التقديرية لمعامل الثبات.

الصيغة (٢١) لكيودر وريتشاردسون :

يتطلب استخدام الصيغة (٢١) تـوافر شرط آخر إلى جانب الشروط الأخرى التي تتطلبها الصيغة (٢٠) ، وهذا الشرط هو تسـاوى جميع مفردات الاختبارات والمقاييس في درجة صعوبتها أو على الأقل يتراوح متـوسط درجة صعوبة جميع المفردات ٠,٥٠ ولكن نظراً لـصعوبة تحـقق هذا الشرط في كـثير من الاختـبارات أو المقـاييس ، فإن الصيغة (٢٠) ربما تكون أكثر ملاءمة لهذه الاختبارات.

وهذه الصيغة مشتقة جبريًا من الصيغة (٢٠) ، وهي كالتالي:

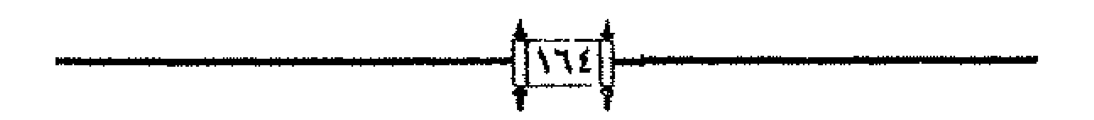
$$\text{معامل التجانس} = \frac{ن}{ن - ١} \left[١ - \frac{س (ن - س)}{ن \times ع^{٢}} \right] \quad ..(٤-١١)$$

حيث (س) ترمز إلى متوسط الدرجات الكلية في الاختبار.

أما بقية الرموز فقد سبق تعريفها في الصيغة (٢٠).

ويتطلب استـخدام هذه الصيغة إيجـاد المتوسط والانحراف المعـيارى للدرجات الكلية فقط، والتعويض في الصيغة (٤-١١).

ومن الجدير بالذكـر أن الصيغة (٢٠) هي الصيـغة الأكثر شيوعًا واسـتخدامًا في القياس التربوى والنفسي؛ وذلك لأن الصيغة (٢١) تعطى قيمًا تقديرية لمعامل التجانس أقل من حقيقـتها، وبخاصة إذا استُخـدمت دون مراعاة للشروط التي تتطلبـها، كما أن الصيغة (٢٠) تتميز بسهولة عملياتها الحسابية.



وعمومًا، فإن معامل التجانس يتعلق بتقييم مدى قياس المفردات المختلفة التى يشتمل عليها الاختبار لسمة أحادية البعد. فإذا كان الاختبار يقيس أكثر من سمة واحدة، فإن معامل التجانس يقلل من القيمة التقديرية لثبات درجات الاختبار.

معامل (α) لكرونباخ :

يتضح مما سبق أن كلا من صيغتى كيودر وريتشاردسون تستخدم في الاختبارات التى تشتمل على مفردات ثنائية الدرجة، غير أنه إذا كان الاهتمام منصبًا على بناء مقياس متدرج الميزان مثل موازين التقدير، أو استبيانات قياس الاتجاهات واستطلاع الرأى أو بعض مقاييس الشخصية التى يستجيب الفرد لعبارات المقياس على ميزان ثلاثى أو خماسى التدريج مثل (موافق جدًا = ٥ ، موافق = ٤ ، غير متأكد = ٣ ، غير موافق = ٢ ، غير موافق على الإطلاق = ١) فهنا لا نستطيع اعتبار إحدى الاستجابات صحيحة والأخرى خطأ وإنما تقع الاستجابات على متصل يتراوح بين موافق جدًا وغير موافق على الإطلاق.

وكذلك فى اختبارات التحصيل التى تشتمل على أسئلة مقال يتم تقدير درجاتها بحسب نوعية الإجابة التى ربما تكون من مستويات متعددة، فعندئذ لا تصلح أى من صيغتى كيودر وريتشاردسون فى تقدير ثبات درجات مثل هذه المقاييس والاختبارات.

وقد تمكن Cronbach من اشتقاق صيغة عامة من الصيغة (٢٠) السابقة لتقدير ثبات درجات أنواع الاختبارات والمقاييس المختلفة. وتؤدى هذه الطريقة إلى معامل اتساق داخلى لبنية الاختبار أو المقياس ويسمى أيضًا معامل التجانس ، غير أنه أطلق على هذه الصيغة معامل (α) ، وهى كالتالى :

$$\text{معامل } (\alpha) = \frac{\text{ن}}{\text{ن} - ١} \left[١ - \frac{\text{مجـ } \text{ع}^{٢}_{\text{ن}}}{\text{ع}^{٢}} \right] \quad \ldots (١٢\text{-}٤)$$

حيث ($\text{ع}^{٢}_{\text{ن}}$) ترمز إلي تباين درجات كل مفردة من مفردات الاختبار.

(مجـ $\text{ع}^{٢}_{\text{ن}}$) ترمز إلى مجموع تباين درجات جميع المفردات.

(ن) ترمز إلى العدد الكلى لمفردات الاختبار.

والصيغة (٤-١٢) تختلف عن الصيغة (٤ ـ ١٠) فى بسط الكسر الثانى فقط حيث حلت (مجـ $\text{ع}^{٢}_{\text{ن}}$) محل (مجـ س ص) ،وكلاهما يرمز إلى تباين الدرجات الكلية للاختبار بصيغتين مختلفتين .

ومن الجدير بالذكر أن قيمة معامل (α) تساوى متوسط القيم التقديرية لمعامل ثبات كل من نصفى الاختبار لجميع طرق التجزئة النصفية الممكنة، كما هو الحال فى معامل التجانس لكيودر وريتشاردسون . وبعبارة أخرى فإن معامل (α) يعطى قيمة تقديرية للارتباط بين عينتين عشوائيتين من المفردات المستمدة من نطاق شامل للمفردات التى تماثل مفردات كل من العينتين.

وقد وجد كرونباك Cronbach أن هذا المعامل يعد مؤشراً للتكافؤ أى يعطى قيماً تقديرية جيدة لمعامل التكافؤ، إلى جانب الاتساق الداخلى أو التجانس. ومن الملاحظ أن معامل (α) ومعامل التجانس لكيودر وريتشاردسون يتأثران بطول الاختبار أى عدد مفرداته ومدى تجانس هذه المفردات. ويعطى معامل (α) الحد الأدنى للقيمة التقديرية لمعامل ثبات درجات الاختبارات ، أى أن قيمة معامل الثبات بعامة لا تقل عن قيمة معامل (α)، فإذا كانت قيمة معامل (α) مرتفعة، فإن هذا يدل بالفعل على ثبات درجات الاختبار، أما إذا كانت منخفضة فربما يدل ذلك على أن الثبات يمكن أن تكون قيمته أكبر من ذلك باستخدام طرق أخرى.

ولتوضيح كيفية تطبيق معامل (α) عندما يكون المقياس متدرج الميزان، نفترض أن لدينا مقياساً للاتجاهات يشتمل على ٦ عبارات على ميزان ثلاثى التدريج (موافق = ٣ ، محايد = ٢ ، غير موافق = ١) طبق على خمسة أفراد ، وكانت درجاتهم كالتالى (جدول ٤ ـ ٢) :

الفرد	١	٢	٣	٤	٥	٦	
١	٢	١	٣	٢	٣	١	١٢
٢	١	٢	٣	٣	٢	٣	١٤
٣	٢	١	٢	٣	١	٢	١١
٤	٢	٣	٣	١	١	١	١١
٥	١	٢	٣	٢	٣	٢	١٣
المجموع	٨	٩	١٤	١١	١٠	٩	٦١
المتوسط س	١,٦٠	١,٨٠	٢,٨٠	٢,٢٠	٢,٠٠	١,٨٠	١٢,٢٠
التباين ع^{2}	٠,٢٤	٠,٥٦	٠,١٦	٠,٥٦	٠,٨٠	٠,٥٦	٦,٨٠

جدول (٤ ـ ٢) يوضح توزيع درجات خمسة أفراد فى مقياس للاتجاهات



الخطوة الأولى : نوجد قيمة تباين درجات كل عبارة، وذلك بطرح الدرجة من متوسط درجات العبارة، وتربيع الناتج وقسمته على عدد الأفراد (٥) ، وذلك بعد تكوين الجدول المبين.

الخطوة الثانية: نوجد قيمة تباين الدرجات الكلية في المقياس وذلك بطرح الدرجة الكلية لكل فرد من المتوسط العام للدرجات، وتربيع الناتج وقسمته على عدد الأفراد.

الخطوة الثالثة : نطبق صيغة معامل (α) لكرونباك (٤ ـ ١٢) كالتالي:

القيمة التقديرية لمعامل (α) =

$$\frac{٦}{١-٦}\left[١ - \frac{٠,٢٤+٠,٥٦+٠,١٦+٠,٥٦+٠,٨٠+٠,٥٦}{٦,٨٠}\right]$$

$$= \frac{٦}{٥}\left[١ - \frac{٢,٨٨}{٦,٨٠}\right]$$

$$= ١,٢ \times ٠,٥٨$$

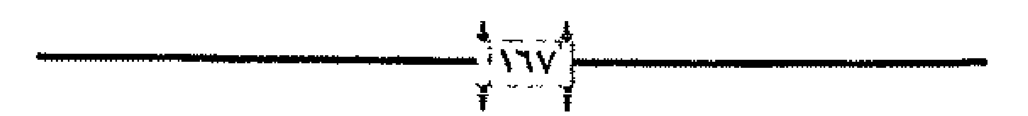
= ٠,٧٠ ويلاحظ أنها قيمة متوسطة .

ثبات تقديرات المحكمين :

يحتاج كثير من الباحثين إلى مؤشر لثبات تقديرات المحكمين وبخاصة فيما يتعلق بأساليب قياس الشخصية الإنسانية التي تعتمد على الملاحظة أو المقابلة الشخصية أو موازين التقدير أو الطرق الإسقاطية التي سوف نوضحها في الفصل الثالث عشر وكذلك في قياس التحصيل الدراسي باستخدام اختبارات تشتمل على أسئلة المقال حيث تعتمد جميع هذه الأساليب على تقديرات مجموعة من المحكمين.

وهنا يبرز التساؤل حول مدى الثقة في هذه التقديرات، أي ما مقدار الخطأ الذي يعزى إلى فردية المحكم أو المحكمين، ويؤثر في الدرجات التي يقدرونها للأفراد.

ولعل أبسط الطرق هو إيجاد قيمة معامل الارتباط بين تقديري محكمين مختلفين فمثلاً يقوم أحد المحكمين بتقدير درجة سؤال مقال معين أو يقوم بملاحظة فرد في موقف معين، ويقوم محكم مستقل آخر بتقدير درجة السؤال نفسه أو يقوم بملاحظة



الفرد نفسه في موقف آخر في ظروف مشابهة،فعندئذ تدل قيمة معامل الارتباط على مدى اتساق تقدير كل منهما أو اتفاقهما.

أما إذا أردنا معرفة مدى اتساق متوسط تقديرات اثنين من المحكمين لمجموعة من الأفراد، فإنه يمكن إيجاد قيمة معامل الارتباط بين تقديرات كل منهما، وتطبيق صيغة سبيرمان وبراون (٤ ـ ٨) التي ذكرناها فيما سبق، حيث نعوض عن ر٢١ بقيمة معامل الارتباط لكي نحصل على القيمة التقديرية لثبات متوسط التقديرات (ر١١).

وفي حالة تعدد المحكمين يمكن إيجاد قيم معامل الارتباط بين تقديراتهم لمجموعة الأفراد واستخدام الصيغة العامة لسبيرمان وبراون (٤ ـ ٧) للحصول على القيمة التقديرية لثبات متوسط تقديراتهم، حيث نعوض عن قيمة (ك) في هذه الصيغة بعدد المحكمين.

ويمكن أيضاً تقدير الثبات عن طريق تحديد النسبة المئوية للاتفاق بين المحكمين أو القائمين بالمقابلة الشخصية إذا كان الهدف مقصوراً على تصنيف الأفراد إلى مجموعتين مثل راسب أو ناجح، متمكن أو غير متمكن، منطوي أو منبسط، وهكذا.

بعض الطرق الأخرى لتقدير الثبات :

هناك بعض الطرق الأخرى التي تستخدم في تحديد تجانس مفردات الاختبار، وتعتمد على أساليب إحصائية متقدمة مثل التحليل العاملي Factor Analysis ،وتحليل التباين Analysis of Variance . فالتحليل العاملي هو أسلوب إحصائي لتحديد أقل عدد من العوامل أو التكوينات الفرضية التي تفسر قيم معاملات الارتباط بين مفردات الاختبار، وبذلك نستطيع باستخدام هذه العوامل تصنيف هذه المفردات في مجموعات متجانسة تقيس كل مجموعة منها سمة معينة. أما إذا بينت نتائج التحليل العاملي وجود عامل مشترك يجمع المفردات فإن هذا يعني أن مفردات الاختبار متجانسة فيما تقيسه.

أما تحليل التباين فهو أسلوب إحصائي آخر يعتمد على تجزئة التباين الكلي لدرجات الاختبار إلى ثلاثة مصادر للتباين ترجع إلى الأفراد والمفردات وتفاعل الأفراد مع المفردات.

ويمكن تقدير تباين الدرجات باستخدام مجموع مربعات انحرافات درجات الأفراد عن المتوسط العام لدرجاتهم، أما تباين الخطأ فيمكن تقديره باستخدام مجموع



المربعات الخاص بتفاعل الأفراد مع المفردات، وبذلك يمكن تقدير ثبات درجات الاختبار باستخدام الصيغة التالية لتحليل التباين التى تنسب إلى هويت Hoyt :

$$\text{معامل الثبات} = \frac{\text{التباين المتعلق بالأفراد} - \text{التباين المتعلق بتفاعل الأفراد مع المفردات}}{\text{التباين المتعلق بالأفراد}} \quad \ldots\ldots\ldots\ldots (١٣ـ٤)$$

والقيمة التقديرية لهذا المعامل تدل على مدى تجانس مفردات الاختبار أو اتساقه الداخلى.

ويمكن للقارئ المهتم باستخدام هاتين الطريقتين الرجوع إلى أحد المصادر الإحصائية التى تناولت التحليل العاملى وتحليل التباين، وقد تناول المؤلف هذان الأسلوبان فى مرجع آخر.

ثبات الفرق بين درجتى اختبارين :

يهتم الباحثون ومستخدمو الاختبارات فى بعض الأحيان بمقارنة درجة فرد فى اختبارين أو أكثر ، فمثلاً ربما نود مقارنة درجة فرد فى مقياس المسئولية الاجتماعية بدرجته فى مقياس تقدير الذات، أو درجته فى اختبار التحصيل فى الرياضيات بدرجته فى اختبار الذكاء، أو درجته فى اختبارين طبق أحدهما قبل عملية التعليم وطبق الآخر بعد انتهائها، فعندئذ يجب أن نتحقق من ثبات الفرق بين كل من الدرجتين لمعرفة تأثير الخطأ العشوائى فى هذا الفرق الذى ربما يؤدى إلى اختلاف الدرجات الملاحظة عن الدرجات الحقيقية. فدرجات كل من الاختبارين تتأثر بعوامل خطأ متباينة، وتتجمع هذه الأخطاء فى درجة الفرق بين الدرجتين.

ولذلك فإن القيمة التقديرية لمعامل ثبات فرق الدرجتين تكون عادة أقل من قيمة ثبات كل درجة على حدة. وهذه القيمة التقديرية تساوى نسبة تباين الفرق بين الدرجتين الملاحظتين إلى نسبة تباين الفرق بين الدرجتين الحقيقيتين، ويمكن إيجاد هذه القيمة باستخدام الصيغة التالية:

$$ر_{\text{الفرق بين الدرجتين}} = \frac{\frac{ر_{١١} + ر_{٢١}}{٢} - ر_{٢١}}{١ - ر_{٢١}} \quad \ldots\ldots (١٤ـ٤)$$

حيث $ر_{١١}$ ترمز إلى معامل ثبات درجات أحد الاختبارين.

، $ر_{٢٢}$ ترمز إلى معامل ثبات درجات الاختبار الآخر.

، $ر_{٢١}$ ترمز إلى معامل الارتباط بين درجات الاختبارين.

أى أن قيمة ثبات درجات كل من الاختبارين على حدة ، وكذلك قيمة معامل الارتباط بينهما تؤثر فى قيمة ثبات الفرق بين الدرجتين.

فإذا كانت قيمة معامل ثبات درجات كل من الاختبارين ٠,٧٥ ، ٠,٨٠ ، وقيمة معامل الارتباط بينهما ٠,٢٥ ، فباستخدام الصيغة (٤ ـ ١٤) نجد أن معامل ثبات الفرق بين درجتى الاختبار

$$= \frac{\frac{٠,٨٠ + ٠,٧٥}{٢} - ٠,٢٥}{١ - ٠,٢٥}$$

$$= \frac{٠,٥٢٥}{٠,٧٥٠} = ٠,٧٠$$

ويلاحظ أن القيمة الناتجة أقل من قيمة معامل ثبات درجات كل اختبار على حدة. وكلما قل الارتباط بين درجات كل من الاختبارين وزاد ثبات كل منهما تزداد قيمة معامل ثبات الفرق بين الدرجتين، والعكس يكون صحيحًا.

فإذا كانت قيمة متوسط ثبات درجات كل من الاختبارين مساوية لقيمة معامل الارتباط بينهما، فإن معامل ثبات الفرق بين درجتيهما يساوى صفرًا، أما إذا كانت قيمة معامل ثبات كل منهما واحدًا صحيحًا فإن قيمة معامل ثبات الفرق بين الدرجتين سوف يكون مساويًا واحدًا صحيحًا أيضًا بغض النظر عن قيمة معامل الارتباط بين درجات كل من الاختبارين. ويمكنك إثبات ذلك بالتعويض فى الصيغة (٤ ـ ١٤) عن :

$ر_{١١}$ = ١ $ر_{٢٢}$ = ١ مهما كانت قيمة $ر_{٢١}$.

وعلى الرغم من أنه يمكننا التحقق من ارتفاع قيمة معامل ثبات درجات اختبارين قبل اختيارهما لمقارنة درجة فرد فى كل منهما ، إلا أننا لا نستطيع أن نقلل قيمة معامل الارتباط بين درجات الاختبارين، وبخاصة إذا كان الاختباران يقيسان نفس المحتوى ويطبق أحدهما قبل عملية التعليم والآخر بعدها.

فهنا يفترض أن درجات الاختبارين تكونان مرتبطتين ارتباطًا مرتفعًا. ولذلك فإن القيمة التقديرية لمعامل ثبات درجات التحسن تكون عادة منخفضة. ولذلك يفضل



استخدام الصيغة (٤ ـ ١٤) في التحقق من ثبات الفرق بين درجتي اختبارين يقيسان سمتين مختلفتين وكلاهما يتميز بثبات مرتفع نسبيًا.

معامل إمكانية التعميم:

يرى كرونباك Cronbach وراجارتنام Rajaratnam وجليزر Glaser أن درجات الاختبار أو أى درجات ملاحظة تعد عينة من نطاق شامل للدرجات الممكنة. وتتباين ظروف جمع الملاحظات وفقًا لأبعاد متعددة.

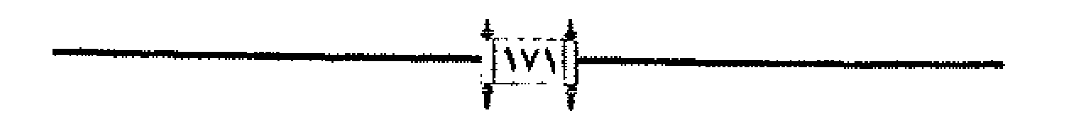
ومثال ذلك تطبيق اختبار معين مرة واحدة، أو تطبيقه وإعادة تطبيقه، واشتماله على عينة ممثلة من المفردات دون بقية العينات المستمدة من نطاق شامل، وتطبيقه فى ظروف معينة دون باقى الظروف.

وهذا التباين فى الأبعاد المتعلقة بأداة القياس وما تشتمل عليه من مفردات، وظروف تطبيقها، وخصائص مجموعة الأفراد التى تطبق عليها تؤدى إلى مشكلات فى عمليات القياس وإمكانية تعميم نتائجها. وقد اقترح كرونباك وزملائه عام ١٩٧٢ نظامًا لتقدير تأثير كل من هذه الأبعاد فى تباين الدرجات الملاحظة ، وكذلك تقدير دقة الدرجة الحقيقية من الدرجات الملاحظة ، وقد أصبح هذا النظام معروفًا باسم نظرية إمكانية التعميم Generalizability Theory.ونظرًا لأن الدرجات الملاحظة تتأثر بجميع هذه الأبعاد، فإن الدرجة الحقيقية تسمى درجة النطاق الشامل Universe Score. والمشكلة الرئيسة فى هذه الحالة تتعلق بإمكانية تعميم ما تدل عليه الدرجات الملاحظة فى ضوء هذه الأبعاد المتعددة.

وبعبارة أخرى ، فإن معامل الثبات فى إطار النظرية الكلاسيكية للاختبارات لا يأخذ بعين الاعتبار تعدد أبعاد عملية القياس، وإنما يركز على أحد هذه الأبعاد.

ولكى يكون تعميم القياس ممكنًا لابد من مراعاة هذه الأبعاد وانعكاس أثرها فى قيمة معامل إمكانية التعميم Generalizability Coefficient.

ونظرًا لأهمية هذه النظرية وإسهاماتها فى توسيع نطاق استخدامات كل من مفهومى الصدق والثبات وتكاملهما، فسوف نلقى عليها مزيدًا من الضوء فى الباب الرابع (الفصل الرابع عشر).



تعقيب عام على طرق تقدير الثبات وتفسير نتائجها :

لعل تعدد طرق تقدير ثبات درجات الاختبارات كما أوضحنا يجعل القارئ يتساءل عن كيفية اختيار إحدى هذه الطرق عند تحققه من ثبات درجات الاختبار أو المقياس الذى يعده أو يستخدمه.

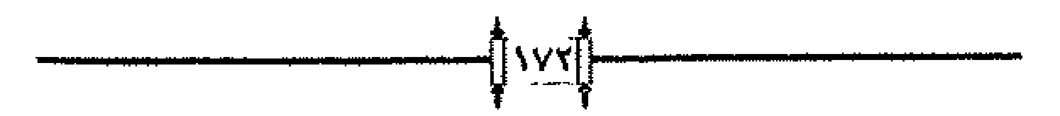
والحقيقة أنه لا توجد إجابة مباشرة أو محددة لهذا التساؤل، إذ إن الاختيار ينبغى أن يستند إلى عوامل متعددة يتعلق بعضها بطبيعة السمة المقاسة، ونوع الاختبار وفيم تستخدم نتائجه، وخصائص عينة المختبرين، والبعض الآخر يتعلق بمصادر الأخطاء العشوائية التى يود أن تؤخذ بعين الاعتبار فى تقدير معامل الثبات، وكذلك الأخطاء التى يرى أنها متسقة أو منتظمة.

فكما رأينا هناك طرق تعتمد على تطبيق الاختبار مرة واحدة، وطرق أخرى تعتمد على تطبيق الاختبار أو صيغ مكافئة له مرتين، واختيار أى من هذه الطرق يعتمد على إمكانات الباحث أو مستخدم الاختبار من حيث توافر صيغتين متكافئتين للاختبار أو تمكنه من إعدادهما، وسهولة تطبيقهما على مجموعة من الأفراد إما فى آن واحد أو بفاصل زمنى معين.

وهذا يتطلب مراعاة ظروف الأفراد وقبولهم إجراء الاختبار عليهم دون تعب أو ملل، كما يتطلب مراعاة ما إذا كان الاختبار أو كل من صيغتيه من نوع اختبارات السرعة أو اختبارات القوة، وكذلك نوع القرار الذى سيتخذ بالنسبة للفرد فى ضوء نتائج الاختبار. فإذا كان القرار يتعلق بالفرد فإن قيمة معامل ثبات درجات الاختبار ينبغى أن تكون مرتفعة بدرجة أكبر مما لو كان القرار متعلقًا بمجموعة من الأفراد.

إذ يرى خبراء القياس أنه ينبغى فى الحالة الأولى ألا تقل هذه القيمة عن ٠,٨٥ كما فى حالة استخدام الاختبارات فى مجالات التشخيص الكلينيكى والتصنيف، ولا تقل فى الحالة الثانية عن ٠,٦٥ ،ومع هذا فإن هاتين القيمتين ليستا مؤكدتين، وإنما تعدان بمثابة مرشد لمستخدم الاختبار. فقرارات الانتقاء تختلف عن قرارات التصنيف أو التشخيص فى تأثيرها على مستقبل الفرد مما يتطلب توافر معلومات أخرى تستند إليها هذه القرارات إلى جانب المعلومات المتعلقة بثبات درجات الاختبارات والمقاييس.

ولذلك فإن قيم معامل الثبات التى يتم تقديرها باستخدام أى من هذه الطرق تختلف فى معناها ودلالتها وذلك لاختلاف مصادر الأخطاء العشوائية التى ينعكس أثرها فى هذه القيم.



فالطريقة التى تعتمد على صيغتين متكافئتين من الاختبار يطبقان فى مرتين متباعدتين تعد أكثر الطرق مراعاة لمصادر الأخطاء العشوائية الناجمة عن التغيرات التى تحدث للأفراد خلال المدة الزمنية الفاصلة بين مرتى التطبيق، والتغيرات الناجمة عن عدم كفاية تمثيل عينة المفردات للنطاق الشامل الذى تقيسه، وكذلك الأخطاء الناجمة عن ظروف تطبيق الاختبار. وتنعكس جميع هذه الأخطاء فى معامل الاستقرار والتكافؤ مما يجعل قيمته التقديرية أقل من قيمة المعاملات الأخرى.

أما طريقة إعادة تطبيق الاختبار نفسه فإنها تراعى فقط الأخطاء العشوائية الناجمة عن التغيرات اليومية التى تطرأ على الأفراد المختبرين، والأخطاء الناجمة عن ظروف التطبيق، مما يجعل القيمة التقديرية لمعامل الاستقرار أقل من قيمة معامل الاستقرار والتكافؤ لعدم تأثر معامل الاستقرار بأخطاء معاينة مفردات الاختبار.

أما طرق الاتساق الداخلى مثل طريقة التجزئة النصفية، وطريقة كيودر ريتشاردسون، ومعامل (α) فإنها تختلف أيضًا فيما بينها من حيث مصادر الأخطاء العشوائية التى تؤخذ بعين الاعتبار فى تقدير قيم معامل الثبات.

فالأخطاء الناجمة عن اختلاف عينة مفردات كل من نصفى الاختبار يؤدى إلى انخفاض قيمة معامل الاتساق الداخلى، فى حين أن هذا المعامل لا يعكس الأخطاء الناجمة عن التغيرات التى تحدث للأفراد أو لظروف التطبيق.

غير أن اعتماد هذه الطرق الثلاث على تطبيق الاختبار مرة واحدة يجعلها مناسبة من الناحية العملية فيما يتعلق بالوقت المخصص لتطبيق الاختبار،وعدم ضرورة توافر صيغ مكافئة للاختبار. ولكن هذه الطرق لا تناسب الاختبارات التى تعتمد على السرعة أو الاختبارات غير متجانسة المحتوى، فكثير من اختبارات الاستعدادات تكون موقوتة وتستخدم فى اتخاذ قرارات متعددة بعيدة الأمد فى المؤسسات التربوية والمهنية.

لذلك ربما يكون من المناسب استخدام طريقة إعادة تطبيق الاختبار فى تقدير ثبات درجات هذه الاختبارات بدلا من التحقق من اتساق درجات كل من نصفى الاختبار. ولكن إذا كان الاختبار يقيس استعدادات متعددة، فإنه ينبغى تقدير ثبات درجات الاختبارات الفرعية التى يشتمل عليها الاختبار.

وإذا كان من الضرورى استخدام طريقة التجزئة النصفية لهذه الاختبارات فيمكن إجراء هذه التجزئة بالنسبة لزمن إجابة مفردات كل من نصفى الاختبار بعد ترتيب المفردات بحسب درجة صعوبتها. أما اختبارات التحصيل الدراسى فإنه يستدل من درجات الطلاب فى عينة من المفردات التى تشتمل عليها على درجة تمكنهم من نطاق سلوكى لمعارف ومهارات معينة. لذلك فإنه من المناسب فى هذه الحالة تقدير ثبات الدرجات باستخدام أى من طرق الاتساق الداخلى. أما إذا اشتمل الاختبار على أسئلة مقال فإن معامل (α) يكون هو المعامل المناسب، والقيم التى نحصل عليها من هذا

المعامل تعد أدنى قيم معاملات الثبات التى يمكن الحصول عليها من المعاملات الأخرى. ونظراً لأننا لا نتخذ قرارات بعيدة الأمد عن تحصيل الطلاب فى منهج دراسى معين، فإن معامل الاستقرار لا يناسب هذه الاختبارات.

فى حين أن هذا المعامل يكون أكثر ملاءمة لمقاييس الجوانب الوجدانية والشخصية نظراً لأن السمات غير المعرفية أكثر استقراراً على المدى البعيد.

كما يفضل استخدام معامل (α) إذا كان الهدف تقدير الاتساق الداخلى لمفردات الاختبارات التى تقيس هذه السمات نظراً لأنها تشتمل على موازين متدرجة. وفى جميع الأحوال ينبغى ألا تقتصر أدلة الاختبارات على تقديم مؤشرات الثبات فقط، وإنما يجب أن تقدم المزيد من المعلومات حول خصائص مجتمع المختبرين، والعينة أو العينات المستمدة منه وعدد أفرادها ومدى تمثيلها للمجتمع، وكذلك المتوسط والانحراف المعيارى لدرجات عينة الأفراد، التى استندت إليها قيمة معامل الثبات. كما ينبغى أن تزود مستخدم الاختبار بقيم الأخطاء المعيارية للقياس التى سوف نوضحها بعد قليل.

والخلاصة أنه يمكن ترتيب طرق حساب الثبات ترتيبًا تنازليًا بحسب حجم القيم التقديرية التى نحصل عليها باستخدامها، وهى كالتالى :

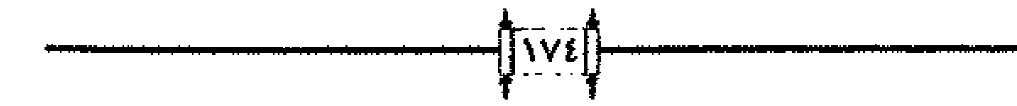
التجزئة النصفية ، كيودر ريتشاردسون ، إعادة تطبيق الاختبار ، الصيغ المتكافئة، معامل (α) .

لذلك ينبغى توقُّع اختلاف هذه القيم باختلاف الطريقة المستخدمة فى موقف معين ولعينة أفراد معينة. فهذه القيم ليست مطلقة، وإنما هى قيم نسبية حيث إنها مرتبطة بكل من الموقف وطبيعة العينة المستخدمة. كما أن هذه القيم تزودنا بتقدير Estimate لدرجة الاتساق فى درجات الاختبار، ولكنها لا توضح أسباب عدم اتساق الدرجات.

الخطأ المعيارى للقياس :

تبين لنا مما سبق أن درجة الفرد الملاحظة فى الاختبار تختلف فى غالبية الأحيان عن درجته الحقيقية نظراً لتأثر الدرجة الملاحظة بمصادر أخطاء متعددة.

لذلك ذكرنا فى مستهل هذا الفصل أن الدرجة الملاحظة تنقسم إلى جزءين هما: الدرجة الحقيقية ودرجة الأخطاء العشوائية. فإذا افترضنا أننا استطعنا تحديد درجة الأخطاء العشوائية التى أثرت فى الدرجة الملاحظة لكل فرد من الأفراد المختبرين، فإنه يمكن إيجاد الانحراف المعيارى لدرجات الخطأ.



والقيمة الناتجة تسمى الخطأ المعيارى للقياس Standard Error of Measurement، ولكننا لا نستطيع فى واقع الأمر إيجاد درجة الخطأ لكل فرد من أفراد المجموعة إلا إذا أعيد تطبيق الاختبار على الفرد نفسه عدداً كبيراً من المرات ، وهو أمر غير ممكن ؛ لذلك فإننا لا نستطيع إيجاد قيمة الخطأ المعيارى للقياس، ولكننا نستطيع تقدير هذه القيمة إذا علمنا قيمة الانحراف المعيارى للدرجات الملاحظة، وكذلك قيمة معامل ثبات درجات الاختبار، وذلك باستخدام صيغة رياضية يمكن اشتقاقها مباشرة من الصيغة (٤ ـ ١). كالتالى:

$$\frac{\text{ع}^{٢}_{\text{ح}}}{\text{ع}^{٢}_{\text{م}}} = ١ - \frac{\text{ع}^{٢}_{\text{خ}}}{\text{ع}^{٢}_{\text{م}}}$$

لكن الكسر الأول يدل على معامل ثبات درجات الاختبار ($\text{ر}_{١١}$) كما أوضحنا من قبل. أى أن:

$$\text{ر}_{١١} = ١ - \frac{\text{ع}^{٢}_{\text{خ}}}{\text{ع}^{٢}_{\text{م}}}$$

أى أن $\text{ر}_{١١} \times \text{ع}^{٢}_{\text{م}} = \text{ع}^{٢}_{\text{م}} - \text{ع}^{٢}_{\text{خ}}$

$$\text{ع}^{٢}_{\text{خ}} = \text{ع}^{٢}_{\text{م}} - \text{ر}_{١١} \times \text{ع}^{٢}_{\text{م}}$$

$$= \text{ع}^{٢}_{\text{م}} \left(١ - \text{ر}_{١١} \right)$$

وباستخراج الجذر التربيعى لكل من الطرفين نجد أن :

$$\text{ع}_{\text{خ}} = \text{ع}_{\text{م}} \sqrt{١ - \text{ر}_{١١}} \quad ... \quad (٤ ـ ١٥)$$

أى أن الخطأ المعيارى للقياس = حاصل ضرب الانحراف المعيارى للدرجات الملاحظة في الجذر التربيعى لمعامل الثبات مطروحًا منه الواحد الصحيح.

فإذا افترضنا أن قيمة معامل ثبات درجات الاختبار ($\text{ر}_{١١}$) تساوى ٠,٨٠ ، وقيمة الانحراف المعيارى للدرجات الملاحظة ٦,٣ ، فإن الخطأ المعيارى للقياس يساوى

$$٦,٣ \sqrt{١ - ٠,٨٠} = ٢,٨٢.$$

ويلاحظ أن الخطأ المعيارى للقياس يساوى صفراً، إذا كانت قيمة معامل الثبات مساوية للواحد الصحيح، ولكنه يصبح مساوياً للانحراف المعيارى للدرجات الملاحظة إذا كانت قيمة معامل الثبات مساوية صفراً.

وهذا يعنى أنه إذا كانت درجات الاختبار غير متسقة على الإطلاق، فإن تشتت الدرجات الملاحظة يعزى إلى محض الصدفة.

ونظراً لأننا لا نستطيع كما ذكرنا إعادة تطبيق الاختبار على الفرد عدداً كبيراً من المرات، فإن الصيغة (٤ ـ ١٥) تعطى قيمة تقديرية للخطأ المعيارى للقياس مستمدة من أفراد المجموعة ككل. غير أن هذه الصيغة يمكن أن تدل أيضاً على الانحراف المعيارى للدرجات الملاحظة للفرد عن درجته الحقيقية لو أن الاختبار أعيد تطبيقه عليه عدداً كبيراً من المرات نظراً لتباين درجات الفرد من تطبيق إلى آخر نتيجة عوامل الخطأ، وبذلك تتذبذب درجاته حول درجته الحقيقية. فإذا افترضنا أن الأخطاء التى تؤثر فى هذه الدرجات كانت عشوائية، فإنها تعمل أحياناً على زيادة درجته الملاحظة، أى تكون درجة الخطأ موجبة، وأحياناً أخرى تعمل على خفض هذه الدرجة، أى تكون درجة الخطأ سالبة.

وبذلك يكون متوسط درجات الخطأ صفراً ، ودرجته الملاحظة تساوى درجته الحقيقية فى الاختبار.

وقد سبق أن أوضحنا فى الفصل الثالث أن المتغير العشوائى يخضع للتوزيع الاعتدالى؛ لذلك فإن الدرجات الملاحظة فى الاختبار تخضع أيضاً للتوزيع الاعتدالى حول الدرجة الحقيقية وذلك لتأثرها بالأخطاء العشوائية. ونظراً لأن درجة الفرد الحقيقية فى الاختبار لا تتغير مهما أعدنا تطبيق الاختبار عليه، فإن تباين هذه الدرجة = صفر. وبالعودة إلى الصيغة (٤ ـ ١) وهى:

$$ع^2_م = ع^2_ح + ع^2_خ$$

والتعويض عن قيمة تباين الدرجة الحقيقية $ع^2_ح$ = صفر.

فإن $ع^2_م = ع^2_خ$ ..... (٤ ـ ١٦).

أى أن تباين درجات الفرد الملاحظة فى التطبيقات المتكررة للاختبار يساوى تباين درجات الخطأ. أما بالنسبة للمجموعة ككل إذا اختبرت مرة واحدة، يكون تباين درجات الخطأ أقل من تباين الدرجات الملاحظة نظراً لأن درجاتهم الحقيقية سوف

تكون مختلفة على الرغم من أن تباين الدرجات الحقيقية لكل منهم على حدة تساوى صفر.

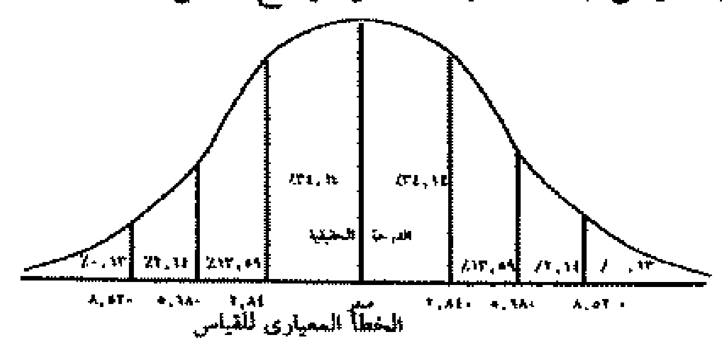
لذلك فإن الخطأ المعيارى للقياس يتم تقدير قيمته من درجات مجموعة من المختبرين، مع افتراض أن هذا الخطأ المعيارى متساو لكل فرد. غير أن هذا الافتراض يبدو غير واقعى وبخاصة إذا كانت مجموعة الأفراد غير متجانسة. ولذلك فإن بعض الاختبارات الجيدة المنشورة تُقَدِّر أخطاء معيارية مختلفة للمجموعات المتباينة من الأفراد سواء فى النوع أو العمر أو الصف الدراسى على متصل السمة التى يقيسها الاختبار. كما أن التطورات المعاصرة التى حدثت فى نظرية الاختبارات أكدت ضرورة تقدير الخطأ المعيارى لكل مستوى معين من مستويات السمة المقاسة، ولعل أهم هذه التطورات ما أصبح معروفًا فى أوساط القياس التربوى والنفسى بنظرية الاستجابة للمفردة الاختبارية Item Response Theory (IRT) التى سوف نلقى الضوء عليها فى الباب الرابع (الفصل الرابع عشر).

والخطأ المعيارى للقياس يفيد فى تقدير مدى تشتت الدرجات الملاحظة حول الدرجة الحقيقية الافتراضية فى اختبار ما، كما يفيد فى تقدير قيمة الخطأ العشوائى الذى يؤدى إلى تذبذب الدرجة الملاحظة للفرد حول درجته الحقيقية.

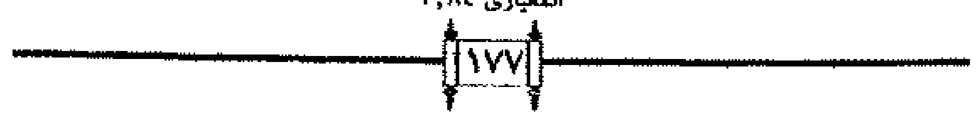
تفسير الخطأ المعيارى للقياس :

لتوضيح كيفية تفسير قيمة الخطأ المعياري للقياس نعود للمثال السابق المتعلق بالاختبار الذى قيمة معامل ثبات درجاته ٠,٨٠ ، والانحراف المعيارى للدرجات الملاحظة للفرد ٦,٣ حيث وجدنا أن القيمة التقديرية للخطأ المعيارى للقياس باستخدام الصيغة (٤ ـ ١٥) مساوية ٢,٨٢.

فالخطوة الأولى : نرسم المنحنى الاعتدالى ونحدد على المحور الأفقى قيم الخطأ المعيارى للقياس ومضاعفاتها كما هو موضح بشكل (٤ ـ ٩) التالى:



شكل (٤ـ٩) يوضح المنحنى الاعتدالى لتوزيع درجات الخطأ التى متوسطها صفر، وانحرافها المعيارى ٢,٨٤



الخطوة الثانية : نحدد مدى الدرجات Score Bands التى تنحصر بين وحدة خطأ معيارى واحد إلى يمين المتوسط وإلى يساره أى + ١ ع خ، وكذلك بين وحدتين وثلاث وحدات أى بين + ٢ع خ ، وبين + ٣ ع خ.

وبذلك يكون مدى الدرجات كالتالى : + ٢,٨٤ ، + ٤,٦٨ ، + ٨,٥٢ .

الخطوة الثالثة : نحدد النسبة المئوية من المساحة تحت المنحنى الاعتدالى المحصورة بين وحدات الخطأ المعيارى المذكورة فى الخطوة الثانية. وهذه النسب موضحة فى شكل (٤ ـ ٩) فنجد مثلا أن :

٦٨ ٪ تقريبًا من المساحة تنحصر بين + ١ ع خ أى بين + ٢,٨٤.

٩٥٪ تقريبًا من المساحة تنح صر بين + ٢ ع خ أى بين + ٤,٦٨ وهكذا.

وهذا يعنى أن الدرجة التى يحصل عليها الفرد فى الاختبار سوف تقع بين +٢,٨٤ حول درجته الحقيقية ٦٨٪ من المرات تقريبًا.

فالدرجة الحقيقية ثابتة، والدرجة الملاحظة تتذبذب حولها، فإذا حصل فرد فى الاختبار على الدرجة ١٢ مثلا فإنه يمكننا أن نستدل على أن درجته الحقيقية سوف تقع بين ١٢ - ٢,٨٤ ، ١٢ + ٢,٨٤ ، أى بين ٩,١٦ ، ١٤,٨٤ ، وثقتنا فى هذا الاستدلال ٦٨٪ تقريبًا .

كذلك يمكننا أن نستدل بدرجة ثقة ٩٥٪ تقريبًا على أن درجته الحقيقية سوف تقع بين ١٢ - ٥,٦٨ ، ١٢ + ٥,٦٨ أى بين ٦,٣٢ ، ١٧,٦٨ .

ويلاحظ أن المدى الذى يحتوى الدرجات الحقيقية يزداد بزيادة درجة ثقتنا، غير أنه يكتفى عادة بالمدى + ١ ع خ فى تفسير القيم التقديرية للخطأ المعيارى للقياس. ولكن إذا أردنا مقارنة الدرجات التى يحصل عليها فرد فى اختبارات متعددة مثل الاختبارات الفرعية التى تشتمل عليها مقاييس الشخصية أو بطاريات الاستعداد العقلى والتى تمثل درجاتها بيانيا فى صفحة نفسية Profile ، فإنه يفضل فى هذه الحالة رسم شريط عريض (مدى) Band حول الدرجة الملاحظة فى كل اختبار بحيث تكون سعته + ٢ ع خ فلكى يكون الفرق بين درجتى الفرد فى اختبارين مختلفين دالا يجب أن يكون أكبر من ضعف الخطأ المعيارى للقياس لأى من الاختبارين.

ويرجع ذلك إلى أن الخطأ المعيارى للفرق بين درجتى الاختبارين أكبر من الخطأ المعيارى لكل اختبار على حدة كما أوضحنا فيما سبق. ولكن ينبغى ملاحظة أن إجراء هذه المقارنة يتطلب أن تكون وحدات القياس المستخدمة فى هذه الاختبارات موحدة نظرًا لأن الخطأ المعيارى للقياس يعبِّر عنه بنفس وحدات الانحراف المعيارى. فإذا

كانت قيمة الانحراف المعيارى تم حسابها من درجات خام، فإن القيمة التقديرية للخطأ المعيارى للقياس ستكون درجات خام أيضًا، وإذا تم حساب قيمة الانحراف المعيارى من نسب ذكاء فإن قيمة الخطأ المعيارى للمقياس سوف يعبر عنها بنسب ذكاء أيضًا، وهكذا.

وهذا يدل على أننا لا نستطيع مقارنة القيم التقديرية للخطأ المعيارى للقياس لاختبارات يختلف نوع معاييرها، لذلك يجب فى هذه الحالة إجراء المقارنة باستخدام القيم التقديرية لمعامل الثبات.

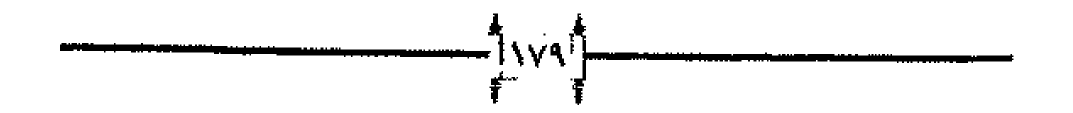
فالقيم التقديرية للخطأ المعيارى للقياس تُعد قيمًا مطلقة وليست نسبية كما هو الحال فى القيم المناظرة لمعامل الثبات، وتُعبِّر عن مدى تكافؤ الدرجات الملاحظة والدرجات الحقيقية للفرد فى الاختبار.

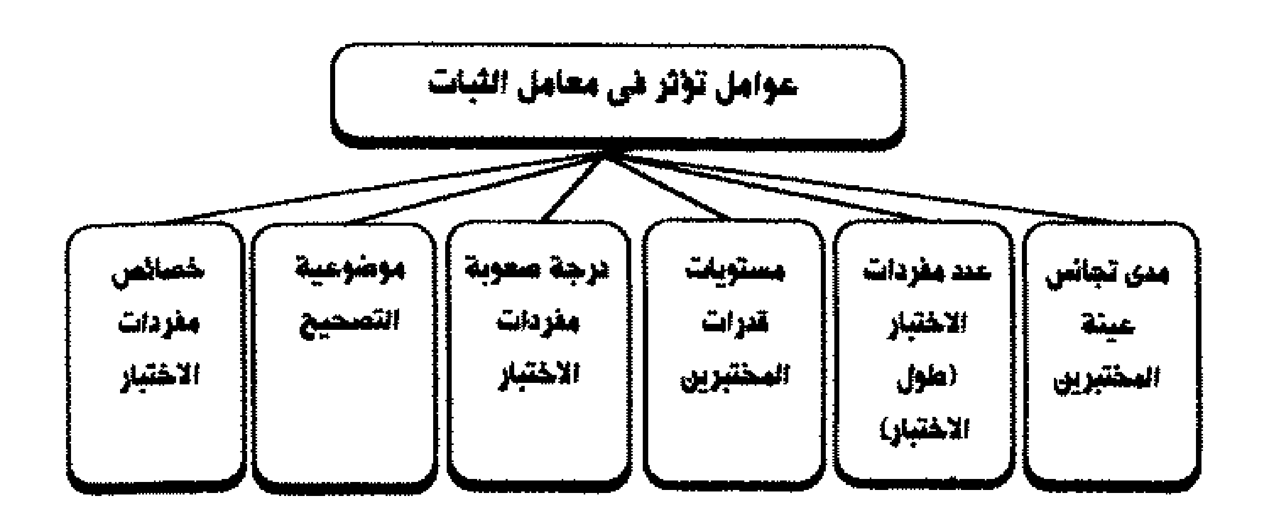
والخلاصة أن الخطأ المعيارى للقياس يعد عاملاً أساسيًا فى تقرير نتائج الاختبارات والمقاييس وتفسيرها، وهو مرتبطًا ارتباطًا وثيقًا بمفهوم الثبات ولكنه أيسر فهمًا لمستخدمى الاختبارات والمختبرين، ولا يتأثر كثيرًا بتشتت الدرجات على عكس الثبات الذى يعتمد على التشتت.

لذلك تهتم أدلة الاختبارات بتحديد القيم التقديرية لهذا الخطأ لكى يتمكن مستخدم الاختبار من تفسير الدرجات تفسيرًا احتماليًا. فإذا حصل فرد فى اختبار ما على الدرجة ٦٥ فإن هذه الدرجة ليست نقطة محددة وإنما يمكن أن تتذبذب فى مدى يتراوح بين قيم معينة على متصل الدرجات تحددها القيمة التقديرية للخطأ المعيارى للقياس ودرجة الثقة المطلوبة. فبدلا من القول بأن الفرد حصل على الدرجة ٦٥ نقول مثلاً: بأننا نثق ٦٨٪ أن درجته الحقيقية تقع بين ٦٣ ، ٦٦ إذا ما أعيد تطبيق الاختبار عليه عددًا كبيرًا من المرات.

عوامل أخرى تؤثر فى قيم معامل الثبات:

سبق أن ذكرنا أن القيم التقديرية لمعامل الثبات تتأثر بمصادر متعددة للأخطاء العشوائية، غير أن هناك العديد من العوامل الأخرى التى تؤثر فى هذه القيم، وهذه العوامل تهم القائمين ببناء الاختبارات والمقاييس أكثر مما تهم مستخدميها، ومع هذا فإن معرفة مستخدمى الاختبارات لهذه العوامل ومراعاتها يساعدهم فى انتقاء اختبارات تتميز درجاتها بالثبات، كما يساعدهم فى تفسير قيم معامل الثبات تفسيرًا مستنيرًا. والشكل التخطيطى (٤ ـ ١٠) التالى يوضح بعض هذه العوامل:





شكل (٤ ـ ١٠) يوضح بعض العوامل المؤثرة في قيم معامل الثبات

وسوف نوضح فيما يلي العوامل الموضحة في شكل (٤ ـ ١٠)

مدى تجانس عينة المختبرين :

تعتمد القيمة التقديرية لمعامل الثبات اعتماداً كبيراً على مدى الفروق بين الأفراد المختبرين، فكلما زادت هذه الفروق ازداد تباين الدرجات الحقيقية للأفراد وبالتالي تزداد قيمة معامل الثبات. فالثبات كما أوضحنا يشير إلى اتساق قياس الفروق الفردية الحقيقية، فكلما زادت هذه الفروق يسهل الحصول على قياسات متسقة إذا تكررت عملية القياس على مجموعة الأفراد، فإذا كانت مجموعة الأفراد متجانسة في القدرة أو السمة التي يقيسها الاختبار فإن تباين الدرجات الحقيقية يقل وبالتالي تنخفض قيمة معامل الثبات أو ربما تصبح مساوية صفراً. ولهذه الخاصية تطبيقاتها الميدانية عند استخدام درجات الاختبارات والمقاييس في اتخاذ قرارات تتعلق بالطلاب أو العاملين الذين تكون قدراتهم متجانسة. فعندئذ ينبغي مراعاة انتقاء اختبارات اعتمد تقدير ثبات درجاتها على عينة من الأفراد مستوى قدراتها في نطاق مستوى المجموعة التي نهتم بدراستها.

فإذا كانت قيمة معامل الثبات المذكورة في دليل الاختبار تم تقديرها من مجموعة من تلاميذ الصفوف الثلاثة الأخيرة من مرحلة التعليم الأساسي، ولكننا طبقنا الاختبار بعد ذلك على تلاميذ صف واحد فقط، فإننا نتوقع أن تقل القيمة التقديرية لمعامل ثبات الدرجات عما هو مذكور في دليل الاختبار نظراً لأن تلاميذ هذا الصف أكثر تجانساً من تلاميذ الصفوف الثلاثة.

عدد مفردات الاختبار

سبق أن أوضحنا عند مناقشتنا معامل الاتساق الداخلى أنه كلما زاد عدد مفردات الاختبار أى طول الاختبار كلما زادت قيمة معامل ثبات درجاته. ويرجع ذلك إلى أن زيادة عدد المفردات يسمح للأخطاء العشوائية الموجبة والسالبة أن تتلاشى بعضها بعضاً مما يجعل الدرجة الملاحظة للفرد فى الاختبار تقترب من درجته الحقيقية. وعند مناقشتنا صيغة سبيرمان وبراون Spearman - Brown Formula أوضحنا أنه يمكن تقدير القيمة المتوقعة لمعامل ثبات درجات الاختبار إذا أطلنا الاختبار أو جعلناه أقل طولاً. ولعل جدول (٤ ـ ١) يوضح كيف أن زيادة عدد مفردات الاختبار تؤدى إلى زيادة القيمة المتوقعة لمعامل الثبات.

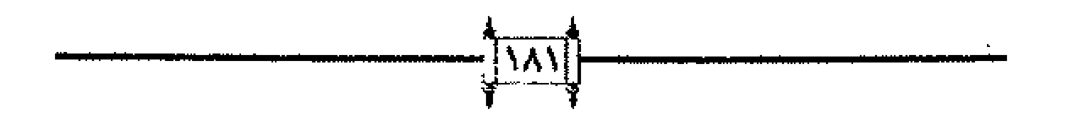
مستويات قدرات المختبرين:

يؤثر مستوى قدرات المختبرين أيضًا فى قيم معامل الثبات الذى نحصل عليه، فتباين درجات الخطأ يزيد بالنسبة للمجموعة منخفضة الدرجات حيث تلعب عوامل التخمين والصدفة دورًا أكبر فى درجاتهم.

لذلك لا يجوز استخدام درجات مجموعة من مستوى قدرة معين للتنبؤ بثبات درجات الاختبار إذا طبق على مجموعة من مستوى آخر. فمثلاً إذا أردنا استخدام اختبار معين لانتقاء الأفراد فى إحدى المهن الدنيا، أو أفراد مبتدئة فى إحدى المهن يجب ألا يتأثر القرار بالبيانات المتعلقة بالثبات المستمدة من أفراد مهنة عليا أو أفراد من ذوى مهارات متقدمة. فالاختبار عندئذ تتباين درجة دقته فى قياس هذه المستويات المختلفة. وهذا يتطلب مراجعة دليل الاختبار بعناية لتعرف مدى تأثر درجات الاختبار بتباين مستوى متوسط قدرات المختبرين.

درجة صعوبة مفردات الاختبار:

تؤثر درجة صعوبة مفردات الاختبار فى قيم معامل الثبات، فإذا كانت المفردات غاية فى السهولة أو الصعوبة، فإنه لا نستطيع باستخدامها قياس الفروق الفردية. ففى الحالة الأولى يستطيع كل فرد أن يجيب إجابة صحيحة عن جميع المفردات، والعكس فى الحالة الثانية، وبذلك يكون توزيع الدرجات منتظمًا فى الحالتين. فدرجة صعوبة الاختبار ككل هى متوسط توزيع درجات الاختبار مقسومًا على عدد مفرداته، فكلما



اقتربت قيمة هذا المتوسط من عدد مفردات الاختبار كان الاختبار أكثر سهولة، وإذا تساوى المتوسط مع عدد المفردات كان الاختبار غاية فى السهولة، وعندئذ يستطيع كل فرد أن يجيب إجابة صحيحة عن جميع مفرداته ويصبح تباين الدرجات صفرًا، وبالتالى تكون قيمة معامل الثبات صفرًا أيضًا. وبالطبع تختلف الاختبارات فى مدى سهولة أو صعوبة مفرداتها، فالاختبارات مرجعية الجماعة أو المعيار Norm - Referenced Tests التى تصمم لقياس الفروق الفردية تكون مفرداتها عادة متوسطة الصعوبة لإبراز هذه الفروق، أما الاختبارات مرجعية المحك Criterion - Referenced Tests التى تصمم للأغراض التعليمية مثل الاختبارات التحصيلية الصفية، فإنها تهدف لقياس تمكن جميع الطلاب من مجال دراسى معين. لذلك فإن مفرداتها ربما تتميز بالسهولة النسبية لكى تقلل من عوامل التخمين الذى يؤدى إلى الأخطاء العشوائية فى الدرجات التى تسهم بدورها فى خفض قيمة معامل الثبات. ولذلك فإن تقدير ثبات درجات هذا النوع من الاختبارات باستخدام الطرق السابقة التى تعتمد على معامل ارتباط بيرسون لا تكون مناسبة، وسوف نشير إلى طرق أخرى تناسب الاختبارات مرجعية المحك فى الفصل الثامن.

موضوعية التصحيح:

سبق أن ناقشنا مفهوم الموضوعية Objectivity ،وتبين لنا أنه كلما تأثر تصحيح الاختبار أو تقدير درجاته بعوامل ذاتية أو عوامل التحيز انخفضت قيمة معامل ثبات الدرجات .

فتصحيح الاختبارات التى تشتمل على مفردات مثل الاختيار من متعدد، أو الصواب أو الخطأ، أو الإكمال يكون عادة موضوعيًا سواء أجرى يدويًا أو آليًا. ولكن المشكلة تبدو واضحة فى تقدير درجة اختبارات المقال وبعض مقاييس الأداء والشخصية، حيث يتضمن التصحيح أحكامًا فردية تتعلق بنوعية الاستجابات، وهذا بدوره يؤثر تأثيرًا بالغًا فى ثبات التقديرات. وقد سبق أن أوضحنا فى هذا الفصل إمكانية التغلب على هذه المشكلة عند مناقشتنا ثبات تقديرات المحكمين.

خصائص مفردات الاختبار:

تؤثر مفردات الاختبار في ثبات درجات الاختبار ككل، فخلو المفردات من الخطأ يعتمد على كيفية بناء هذه المفردات، فبعض المفردات ربما تشتمل على مؤشرات لإجابة مفردات أخرى في الاختبار مما يساعد بعض الأفراد على التخمين بدرجة جيدة مما يعمل على خفض قيمة معامل الثبات.

وكذلك المفردات الغامضة، أو غير محددة الهدف، أو التي تكون صياغتها أو تعليمات إجابتها غير دقيقة، أو غاية في الصعوبة، تؤثر تأثيرًا بالغًا في ثبات درجات الاختبار.

ومن هذا يتضح أن قيم معامل الثبات ليست قيمًا مطلقه وإنما تعد قيمًا تقديرية تؤثر فيها عوامل متعددة ينبغي مراعاتها في تصميم أدوات القياس وعند انتقاء هذه الأدوات واستخدامها وتفسير نتائجها. فالفروض التي تستند إليها طرق تقدير الثبات يصعب تحققها جميعًا في الواقع الميداني، مما يتطلب الحيطة في تفسير قيم معامل الثبات. إذ لا يوجد أسلوب إحصائي يُغني عن الحكم المنطقي والفكر المستنير في تفسير هذه القيم وبخاصة في القياس التربوي والنفسي.

خلاصة

إن الاختبار أو المقياس ينبغي أن تتوافر فيه بعض الخصائص السيكومترية الأساسية من أهمها ثبات درجاته وصدقه. ويشير مفهوم الثبات إلى عدم تأثر الدرجات بالأخطاء غير المنتظمة. وتتعدد مصادر هذه الأخطاء، فبعضها يتعلق بأداة القياس، أو إجراءات تطبيق الاختبار وتصحيحه، وبعضها الآخر يتعلق بالأفراد المختبرين. فإذا انعدمت هذه الأخطاء تصبح الدرجة الملاحظة للفرد في الاختبار مساوية لدرجته الحقيقية التي تعد بمثابة درجة افتراضية. وإذا كان تباين الدرجات الملاحظة مساويًا لتباين الدرجات الحقيقية، فإن تباين الخطأ يكون مساويًا صفرًا، ويصبح الثبات تامًا.

ونظرًا لأننا لا نستطيع قياس التباين الحقيقي بطريقة مباشرة، فإنه يمكننا تقدير قيمة أخطاء الصدفة التي تؤثر في الدرجات باستخدام أساليب مختلفة نحصل منها على قيم تقديرية لمعامل التكافؤ، أو معامل الاستقرار، أو معامل الاستقرار والتكافؤ، أو معامل الاتساق الداخلي، أو معامل التجانس، أو معامل إمكانية التعميم. كما يمكن تقدير ثبات الفرق بين درجتي اختبارين، وثبات تقديرات المحكمين.

كما يمكن إيجاد قيمة الخطأ المعياري للقياس التي تتناسب عكسيًا مع قيمة معامل الثبات، حيث تُحدد مدى معين يحتوي الدرجة الحقيقية للفرد في الاختبار. ويمكن تقليل مصادر الأخطاء العشوائية التي تؤثر في قيم معامل الثبات، وذلك بإطالة الاختبار بقدر معين، ومراعاة الموضوعية في تصحيحه، وكذلك مراعاة تباين الأفراد المختبرين في السمة التي يقيسها الاختبار، والعناية بصياغة وبناء مفرداته، والتحقق من درجة صعوبتها وفقًا لما إذا كان الهدف من الاختبار إبراز الفروق الفردية، أو قياس درجة التمكن في مجال دراسي معين.

الفصل الخامس

الخصائص السيكومترية للاختبارات
ثانيا : صدق الاختبارات

* مقدمة
* جوانب صدق الاختبارات
* صدق المحتوى
* طرق تقدير صدق المحتوى
* الصدق المرتبط بمحك
* الصدق التنبؤى
* طرق تقدير الصدق التنبؤى
* بعض المشكلات المتعلقة بالمحك
* الصدق التلازمي وطرق تقديره
* بعض العوامل التى تؤثر فى الصدق المرتبط بمحك
* صدق التكوين الفرضى
* طبيعة التكوينات الفرضية
* منهجية التحقق من صدق التكوين الفرضى
* أساليب جمع أدلة تتعلق بصدق التكوين الفرضى
* منظور تكاملي لطرق تقدير الصدق والثبات

مقدمة :

تعد جوانب الصدق من أهم خصائص الاختبارات والمقاييس التربوية والنفسية. فصدق الاختبار Test Validity يتعلق بالهدف الذى يبنى الاختبار من أجله، وبالقرار الذى يتخذ استنادًا إلى درجاته، فدرجات الاختبار تستخدم عادة فى التوصل إلى استدلالات معينة، وهنا يبرز التساؤل حول ما يمكن الاستدلال عليه بدرجة عالية من الدقة أو الثقة.

فعلى الرغم من أن ثبات درجات الاختبار يعنى اتساق الدرجات، فإن الصدق يتعلق بدرجة تحققنا من تفسير درجات الاختبار واستخدامها، ولذلك يمكن أن تتميز درجات الاختبار بالثبات أو الاتساق، ولكن لا يعنى ذلك أن الاختبار يكون صادقًا. إذ يمكن أن يُبنى اختبار لقياس المهارات الحسابية ويحصل كل تلميذ على درجات متسقة إذا أعيد تطبيق الاختبار عليه عدة مرات، ومع هذا لا نستطيع الحكم على درجة صدق الاختبار، إذ إن هذا يتطلب معلومات كثيرة تتعلق بمحتوى الاختبار، وعدد مفرداته ونوعها، ومدى تمثيلها للنطاق الشامل للأعداد والعمليات، وما إذا كانت المفردات تشتمل على مسائل كلامية، وغير ذلك مما يؤثر فى أداء التلاميذ فى الاختبار وبالتالى فى تفسير درجاته.

فثبات درجات الاختبار يتأثر فقط بمصادر الأخطاء العشوائية للقياس كما سبق أن أوضحنا، بينما يتأثر صدق الاختبار بكل من مصادر الأخطاء العشوائية والأخطاء المنتظمة أو الثابتة. وهذا يؤكد ما ذكرناه من أن درجات الاختبار يمكن أن تكون على درجة عالية من الثبات دون أن يكون الاختبار صادقًا. ولكن الاختبار لا يمكن أن يتسم بالصدق دون أن تكون درجاته متسقة، أى أن الثبات شرط ضرورى للصدق ولكنه غير كاف، ومع هذا لا نستطيع الفصل بينهما أو اعتبارهما ثنائية. فقد أكد Cronbach أن هذين المفهومين مترابطان ويمكن أن يندرجا تحت اسم «مقاييس قابلية التعميم Measures of Generalizability» التى أشرنا إليها فى الجزء المتعلق بالثبات.

فالفرق الرئيسى بينهما يكمن فى الأبعاد التى يمكن أن تُعمم فى ضوئها نتائج الاختبار أو القياس. كما أكد أن الصدق ليس خاصة تتعلق بالاختبار فى ذاته، وإنما تتعلق بتفسير الدرجات المستمدة من الاختبار. فالاختبار الواحد يمكن أن يستخدم فى أغراض متعددة، فاختبار القراءة يمكن أن يستخدم فى فحص المتقدمين لدورة تدريبية معينة، وفى تشخيص صعوبات القراءة، وفى قياس فاعلية برنامج تعليمى فى اللغة العربية، وغير ذلك. ونظرًا لأن كلا من هذه الاستخدامات يعتمد على تفسيرات مختلفة، فإن الأدلة التى تبرر استخداما معينًا ربما لا تصلح لتبرير استخدام آخر.



فجميع هذه التفسيرات تتميز بقدر معين من الصدق، لذلك يصعب أن نستنتج ببساطة أن الاختبار «صادق Valid». لذلك ليس لكل اختبار مؤشر صدق واحد، وإنما يمكن أن يكون له عدة مؤشرات للصدق بحسب تعدد أغراض استخداماته، ونوع القرارات التي يمكن اتخاذها. فالاختبار الذي يتميز بدرجة معينة من الصدق بالنسبة لغرض معين أو لمجموعة معينة من الأفراد، ربما لا يكون كذلك بالنسبة لغرض آخر أو لمجموعة أخرى.

فالصدق إذن يتعلق بمدى فائدة أداة القياس في اتخاد قرارات تتعلق بغرض أو أغراض معينة. فقرارات الانتقاء مثلا يمكن أن تكون أكثر صوابًا إذا استطاع المعلم تقدير احتمال نجاح الطالب في مقررات مختلفة.

فإذا لم تكن لنتائج أدوات القياس فائدة في اتخاذ قرارات صائبة ربما لأنها لا تقدم معلومات غير دقيقة أو غير متعلقة بمجال القرار، فإن نتائج هذه الأدوات لا تتسم بالصدق مهما بلغت قيمة معامل الثبات.

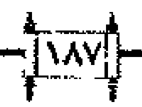
ولكي نحدد مدى ملاءمة الاختبار لاستخدامات معينة ينبغي جمع معلومات مناسبة تتعلق بصدق الاختبار، وتعتمد هذه المعلومات على هدف أو أهداف عملية القياس وليس على نوع الاختبار أو المقياس.

وقد سبق أن ذكرنا في الفصل الأول أن الاختبار يمكن أن يكون بمثابة عينة Sample أو بمثابة مؤشر Sign ،وهذا التمييز يعتمد على درجة تمثيل المفردات للنطاق السلوكي الشامل للسمة المراد قياسها.

فإذا كان الاختبار يشتمل على عينة من المفردات تمثل النطاق السلوكي الشامل تمثيلاً جيداً كما في الاختبارات التحصيلية مثلاً، فإن الاختبار يكون بمثابة عينة Sample، أما إذا كان النطاق السلوكي الشامل غير محدد تحديداً دقيقًا وكاملا كما في بعض مقاييس الشخصية والقدرات، فإن الاختبار أو المقياس يكون بمثابة مؤشر Sign. إذ يمكن أن يستخدم اختبار في الذكاء كمؤشر لبعض الاضطرابات العقلية ، ومثال ذلك اختبار ويكسلر Wechsler للذكاء.

فالذي يطلع على الأدبيات السيكلوجية والتربوية يلاحظ تعدد أنواع الصدق ونمطية استخداماتها بحيث لم يعد ممكنًا التحقق مما يقصد بهذه الأنواع، وما إذا كانت تختلف بالفعل فيما بينها، وربما يرجع ذلك إلى استخدام الباحثين والممارسين لمفهوم الصدق استخدامًا ضيقًا ومحدوداً.

لذلك أصدرت لجنة فنية مشتركة من الجمعية الأمريكية لعلم النفس APA، والجمعية الأمريكية للبحث التربوى AERA ، والمجلس القومي الأمريكي للقياس



التربوى NCME كتيبًا متجددًا يتضمن معايير الاختبارات التربوية والنفسية وأدلتها سبق أن أشرنا إليه في الفصل الثاني، ويؤكد هذا الكتيب أن الاختبارات والمقاييس تستخدم لاتخاذ قرارات وإصدار أحكام متنوعة، وكل قرار أو حكم يتطلب أنواعًا مختلفة من الدراسات للتحقق من الصدق.

ومن بين هذه القرارات والأحكام المتعددة يميِّز الكتيب بين ثلاثة رئيسة منها كالتالي:

(١) تحديد كيفية أداء الفرد في الوقت الحاضر في نطاق شامل لمواقف سلوكية تمثلها مفردات الاختبار تمثيلاً جيدًا. فمثلاً ربما يود مستخدم اختبار معرفة أداء تلميذ في الصف الثاني الابتدائي في الحساب في منتصف العام الدراسي، فالاختبار يكون مشتملاً على عينة ممثلة من المفردات أو الأسئلة المتعلقة بالأهداف السلوكية المرجوة من هذه المادة الدراسية.

(٢) التنبؤ بالأداء المستقبلي أو تقدير مكانة الفرد في أحد المتغيرات ذات الأهمية، ولكنها تختلف عما يقيسه الاختبار. فمثلاً يمكن استخدام اختبار الاستعداد الأكاديمي في التنبؤ بالمعدل العام للطالب في نهاية السنة الأولى بإحدى الكليات، فعندئذ يكون المعدل العام هو المتغير موضع الاهتمام، ويلاحظ أنه يختلف عن متغير الاستعداد الأكاديمي. وكذلك ربما يُستخدم استبيان التوافق في تقدير ما يمكن أن يكشف عنه الفحص السيكولوجي.

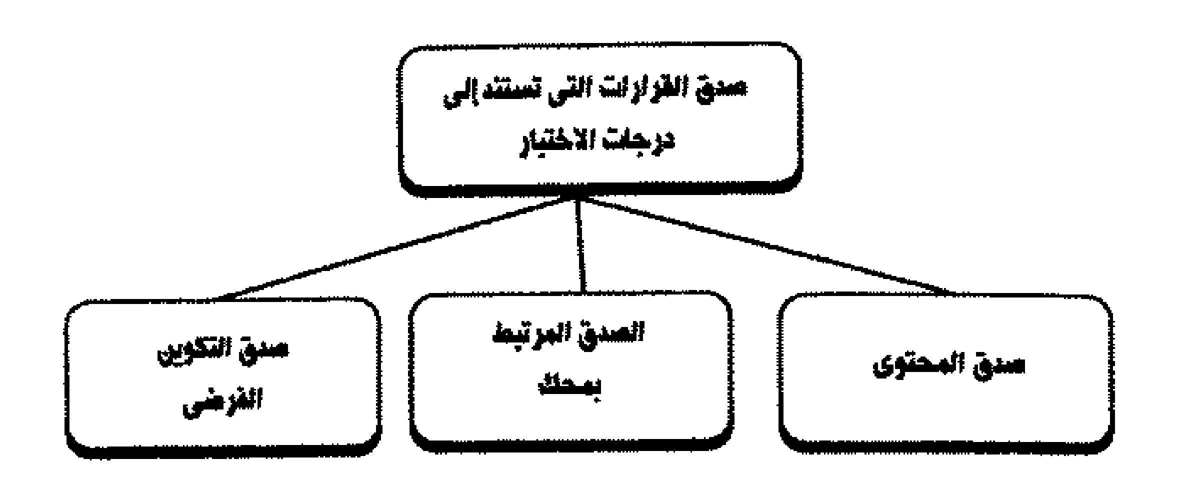
(٣) الاستدلال على درجة تملك الفرد سمة أو خاصة معينة تعد بمثابة تكوين فرضي Construct يبدو أثره في سلوك الفرد أو أدائه. فمثلاً ربما نود معرفة ما إذا كان الفرد يتسم بقدر كبير من سمة معينة مثل الذكاء أو الابتكارية أو الاكتئاب، علما بأن هذه السمات مجردة أي غير منظورة ولا نستطيع ملاحظتها ملاحظة مباشرة.

وربما يكون الهدف من هذا الاستدلال هو تقدير درجة السمة لدى الفرد،أو دراسة الاختبار وتعرف مكوناته، أو دراسة علاقة الاختبار بالاختبارات الأخرى التي تقيس السمة نفسها أو سمات مشابهة، أو لبناء نظرية سيكولوجية.

جوانب صدق الاختبارات:

حددت اللجنة المشتركة ثلاثة جوانب من الصدق تناظر أنواع القرارات الأساسية الثلاثة السابقة، وهي : صدق المحتوى، والصدق المرتبط بمحك، وصدق التكوين الفرضي أو المفهوم كما يوضحه الشكل التخطيطي (٥-١) التالي:





شكل (٥ ـ ١) يوضح الجوانب المختلفة لصدق القرارات التي تستند إلى درجات الاختبارات

ومما هو جدير بالذكر أن هذه الجوانب الثلاثة مترابطة إجرائيًا ومنطقيًا. وقد أكدت اللجنة ذلك، حيث أشارت إلى أنه نادرًا ما يكون أحد هذه الجوانب ذو أهمية بمعزل عن الجانبين الآخرين، فدراسة الاختبار دراسة متعمقة تتضمن عادة معلومات حول الجوانب الثلاثة.

لذلك يرى كثير من علماء القياس مثل كرونباك (Cronbach, 1970) وأنستازي (Anastasi, 1976)، وستانلي (Stanley , 1981) وغيرهم أن الأدلة التي نسترشد بها للتحقق من الصدق وليس الاختبار هي ما ينبغي أن يطلق عليها مفهوم «الصدق».

كما يرون أنه لا ينبغي التفكير في الجوانب الثلاثة كأنواع مختلفة من الصدق، وإنما كأدلة متعددة عن مدى صلاحية الاختبار في ترشيد القرارات. فبقدر ما تكون المعلومات التي يمدنا بها الاختبار مضللة أو غير كافية لإصدار قرارات معينة تكون البيانات المستمدة من هذا الاختبار غير صادقة.

لذلك يؤكد ساكس Sax أنه لا يمكننا أن نبرهن على صدق الاختبار وإنما نستطيع أن نقدم أدلة تتعلق بصلاحيته في غرض أو أغراض معينة.

وعلى الرغم من تكامل الجوانب الثلاثة للصدق إلا أننا سوف نتناول كل منها على حدة بالمناقشة والتحليل لكي يتيسر على القارئ استيعابها.

صدق المحتوى :

يدل صدق المحتوى على مدى تمثيل محتوى الاختبار للنطاق السلوكي الشامل Universe للسمة المراد الاستدلال عليها، إذ يجب أن يكون المحتوى ممثلاً تمثيلاً جيدًا لنطاق المفردات الذي يتم تحديده مسبقًا.

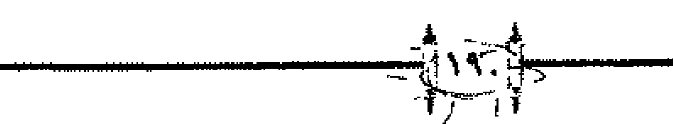
ونقصد بنطاق المفردات: المعارف والمهارات والعمليات التي يتم معاينتها بواسطة مفردات الاختبار. ويمكن إجراء ذلك عن طريق تحليل المحتوى أو النطاق السلوكي بحيث يمثل الاختبار جميع مكونات هذا المحتوى أو هذا النطاق تمثيلاً جيدا. لذلك فإن صدق المحتوى يناسب بدرجة أكبر الاختبارات والمقاييس التربوية، وبخاصة الاختبارات التحصيلية. فإذا أجرى المعلم اختبارًا على تلاميذه لقياس مهاراتهم في استخدام الحاسوب، فإنه يهتم بالتحقق من أن درجات الطلاب في هذا الاختبار تمثل بالفعل المهارات المرجوة. غير أن كثيرًا من العوامل المتعلقة بخصائص مفردات الاختبار ومدى تمثيلها للنطاق السلوكي لهذه المهارات ربما تؤثر في درجاتهم.

وهذه العوامل تؤدي إلى انخفاض صدق الاستدلال على المهارات التي يحددها هذا النطاق. ونظرا لأن الاختبارات التحصيلية واختبارات الكفايات تعد بمثابة عينة Sample ، فإن صدق المحتوى في هذه الحالة يتعلق بأسلوب معاينات المفردات وكفاية العينة المنتقاة التي يشتمل عليها الاختبار.

وهذا يتطلب أن يكون النطاق الشامل للمفردات Items Universe معرفًا تعريفا دقيقًا؛ لأنه يعد بمثابة تعريف إجرائي للسمة المراد قياسها وهي التحصيل في مجال معين. ولعل تحديد الأهداف السلوكية التي تقيسها الاختبارات التحصيلية يعد خطوة ضرورية في هذا الشأن كما سيتضح في الفصل الثامن عند تناولنا أساليب قياس التحصيل.

وقد أشار كتيب اللجنة المشتركة إلى إمكانية التحقق من صدق المحتوى بالنسبة لاختبارات الاستعدادات واستبيانات الشخصية والميول وكذلك مقاييس التوافق والسلوك الاجتماعي، حيث يمكن الاعتماد على عينات من المهارات التي يتضمنها الاستعداد الميكانيكي أو الموسيقي مثلاً، أو عينات من الملاحظات في مواقف منتقاة تتعلق بالميول أو التوافق أو غير ذلك من السمات، ولكن لا ينبغي في هذه الحالات الاقتصار على صدق المحتوى وإنما يجب الاستناد إلى أدلة من جوانب الصدق الأخرى. غير أن بعض علماء النفس لا يتفقون مع هذه اللجنة في هذا الشأن.

إذ ترى أنستازي Anastasi عدم ملاءمة صدق المحتوى لاستخدامات هذه المقاييس، وحجتها في ذلك أن مقياسًا معينًا ربما يقيس وظائف مختلفة لدى



المجموعات المتباينة من الأفراد، مما يجعل تحديد هذه الوظائف من مجرد تأمل أو فحص محتوى المقياس أمراً غير ممكن.

وبوجه عام فإن مدى ملاءمة استخدام صدق المحتوى يعتمد على نوع الاستدلالات أو التفسيرات التي نود التوصل إليها استناداً إلى نتائج الاختبارات والمقاييس.

طرق تقدير صدق المحتوى :

اتضح لنا مما سبق أن صدق المحتوى Content validity يتعلق بمحاولة تحديد مدى كفاية بناء الاختبار بحيث تمثل مفرداته النطاق السلوكي لمهارة أو مجال دراسي معين تمثيلاً جيداً.

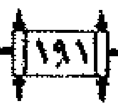
لذلك يتطلب هذا الجانب من الصدق دون غيره من الجوانب أدلة منطقية وليست إحصائية. فتقدير صدق المحتوى يستند إلى منطق مستنير ومهارات تحليلية ومثابرة، حيث إن أدلة هذا الصدق تحتاج إلى مراجعة مستمرة في ضوء التغيرات التي يمكن أن تطرأ على النطاق السلوكي وما يتضمنه من محتوى وعمليات، أو على المناهج الدراسية أو أساليب التعليم.

ولذلك فإن معظم أساليب تقديرصدق المحتوى تعتمد على الأحكام التقييمية لخبراء المواد الدراسية أو المهتمين بتنمية المهارات والكفايات التعليمية والمهنية والفنية، وتتعلق هذه الأحكام بتقدير مدى التناظر بين مفردات الاختبار والنطاق السلوكي الذي تمثله هذه المفردات.

وهذا يتطلب صياغة مكونات النطاق السلوكي من معارف ومهارات وعمليات ومصادر صياغة واضحة ومحددة، وكذلك المكونات الفرعية لهذا النطاق والأهمية النسبية لكل منها، ومحتوى كل مفردة وما تتضمنه من عمليات وإجراءات.

ويمكن إعداد استمارة تشتمل على ميزان تقدير تتضمن الأبعاد الرئيسة المتعلقة بالنطاق السلوكي، مثل محتوى المفردات، ونوع المهارات ومجالها، ومصادر أو مواد ذات أهمية، ونوع المفردات وملاءمتها لكل من المحتوى والمهارة المرجوة.

ويقوم المحكم بتقييم كل مفردة من مفردات الاختبار في ضوء هذه الأبعاد على ميزان التقدير الموضح في الاستمارة. ويفضل الاعتماد على أكثر من محكم للحصول على تقديرات متسقة بدرجة أفضل.



ويمكن التحقق من ذلك بتحليل التقديرات إحصائيًا باستخدام أساليب تحليل التباين أو غيرها من الأساليب التي تزودنا بمؤشرات إحصائية تتعلق بدرجة اتساق التقديرات.

كما يمكن التحقق من صدق المحتوى إمبيريقيا، وذلك بتطبيق الاختبار على مجموعة من الطلاب قبل بدء عملية التعليم، وإعادة تطبيقه عليهم عقب الانتهاء منها، وفحص نتائج الاختبار في المرتين لمعرفة ما إذا كان الاختبار يقيس بالفعل المجال الذي اهتمت به عملية التعليم. وهذا يتطلب أن يكون لدى الطلاب معرفة محدودة بالمجال الذي سوف تتناوله عملية التعليم قبل البدء.

وينبغي التنويه إلى أن بعض الباحثين والمهتمين ببناء الاختبارات ومستخدميها يفترضون أن صدق المحتوى يناظر الصدق الظاهري Face Validity الذي يتحققون منه بإلقاء نظرة سريعة على محتوى الاختبار لتعرف ماذا يقيس، غير أن هذا ربما يكون مفيدًا إلى حد ما في المواقف التي تتطلب انتقاء العاملين للالتحاق بمهن أو وظائف معينة، حيث ربما يكون من المناسب أن يشعر المختبرون بأن الاختبار يبدو أنه يقيس خصائص تتعلق بالمهنة أو الوظيفة المتقدمين لشغلها.

ولكن هناك مواقف أخرى ربما يؤدي الاعتماد فيها على الصدق الظاهري للاختبار أو المقياس إلى نتائج غير مرغوبة وبخاصة في قياس الشخصية باستخدام الاستبيانات، إذ ربما يلجأ الفرد الذي يجيب عن مفردات الاستبيان إلى تزييف استجاباته لشعوره بالتدخل في خصوصياته أو إظهاره بمظهر معين.

لذلك فإن دليل معايير الاختبارات والمقاييس الذي أشرنا إليه استبعد الصدق الظاهري واكتفى بصدق المحتوى.

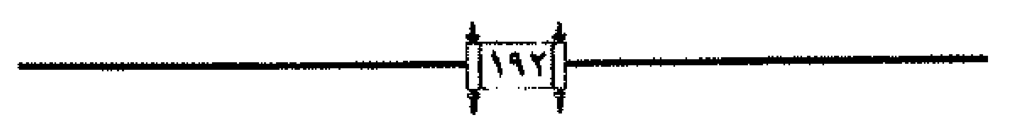
والخلاصة أن تقدير صدق المحتوى يستند إلى فروض ثلاثة صاغها ليتون Lenon كالتالي:

(١) ينبغي أن يكون المجال الذي يُختبر فيه الأفراد محدودًا بنطاق شامل للمفردات الذي تبدو أهميته لهم، ويمكن تعريفه تعريفًا دقيقًا.

(٢) يمكن انتقاء عينة من المفردات من هذا النطاق بطريقة هادفة ومناسبة.

(٣) يمكن تحديد عينة المفردات وأسلوب المعاينات المستخدم وتعريفها بدقة كافية لكي يتمكن مستخدم الاختبار من الحكم على مدى تمثيل عينة المفردات للنطاق السلوكي الشامل الذي تقيسه.

ويلاحظ أن جانب صدق المحتوى شأنه شأن جوانب الصدق الأخرى يتعلق بالغرض من استخدام الاختبار ومجموعة الأفراد المختبرين.



فأدلة صدق المحتوى لمجموعة معينة ربما لا تنطبق بالضرورة على مجموعات أخرى نظرًا لاختلاف العمليات التى تعتمد عليها كل مجموعة فى إجابتهم عن مفردات الاختبار.

لذلك نؤكد مرة أخرى أنه ليس هناك صدق محتوى للاختبار، وإنما صدق محتوى الغرض أو الاستخدام المناسب لمجموعة معينة من الأفراد. وقد وضح كثير من علماء القياس أهمية محاولة تطوير أساليب وإجراءات يمكن باستخدامها تحديد مؤشرات كمية أو مجموعة من المحكات يستند إليها فى تلخيص درجة صدق المحتوى لكى تُيسر عملية التواصل بالمعلومات المتعلقة بهذا الجانب من الصدق.

الصدق المرتبط بمحك:

كثيرًا ما يهتم مستخدمو الاختبارات ومتخذو القرارات بالتنبؤ بنجاح أحد الطلاب فى دراسته الجامعية، أو بفاعلية أداء المعلم بعد تخرجه من كلية التربية، أو بمستوى أداء فرد فى إحدى المهن بعد انتهائه من دورة تدريبية معينة. فالاهتمام فى مثل هذه الحالات ينصب على التنبؤ بالأداء المستقبلى باستخدام درجات اختبارات مناسبة تطبق فى الوقت الحاضر، فتقييم القرارات التنبؤية يعتمد على تقدير الصدق المرتبط بمحك Criterion - Related Validity، أى مدى صلاحية الاختبارات والمقاييس فى أغراض التنبؤ من أجل اتخاذ قرارات عملية.

فالصدق المرتبط بمحك يناسب المواقف التى نود فيها استخدام أداة قياس فى تقدير سلوك معين ذى دلالة، وهذا السلوك خارج عن نطاق الاختبار ذاته ويعد بمثابة للمحك Criterion، ففى الأمثلة السابقة ربما يكون معدل تقديرات الطالب فى نهاية السنة الأولى بإحدى الكليات هو المحك، أو أن نتائج بطاقة تقييم المعلم بعد مرور عام من التحاقه بمهنة التدريس هى المحك، وهكذا.

فالاختبارات التى تُستخدم فى اتخاذ قرارات التنبؤ ليس بالضرورة أن يكون محتواها مماثلاً لمحتوى المقياس المحك، إذ ربما نستطيع أن نتنبأ ببعض الاضطرابات النفسية باستخدام اختبار للذكاء، أو أن نتنبأ بنجاح الطلاب فى دراستهم الجامعية باستخدام اختبارات قبول معينة، أو نتنبأ بالأفراد الذين ينجحون فى المهن التجارية الدعائية باستخدام استبيانات الشخصية.

ففى مثل هذه الحالات كلما كانت التنبؤات دقيقة كان الاختبار أكثر فائدة. ونظرًا لتأثير القرارات المترتبة على نتائج الاختبارات والمقاييس فى مجالات انتقاء الأفراد، فإنه تبدو أهمية توافر الصدق المرتبط بمحك فى هذه القرارات إذا استخدمت الاختبارات والمقاييس لترشيد عمليات الانتقاء.

ومن هذا يتضح أن الصدق المرتبط بمحك يستند إلى الأسلوب الإمبيريقي في دراسة العلاقة بين درجات اختبارات أو مقاييس معينة تعد بمثابة منبئات Predictors ودرجات مقاييس خارجية مستقلة تعد بمثابة محكات Criteria .

غير أن المشكلة الأساسية في تقدير الصدق المرتبط بمحك تتركز في إمكانية الحصول على محكات مناسبة، فجمع بيانات تتعلق بالمحكات تعد من مشكلات القياس التربوي والنفسي التي تتطلب دراسات متعددة. فكما أن الاختبار التنبؤي ينبغي أن يتميز بخصائص معينة، كذلك ينبغي أن يتميز المقياس المحك بخصائص يمكننا من الثقة فيما يترتب عليه من بيانات.

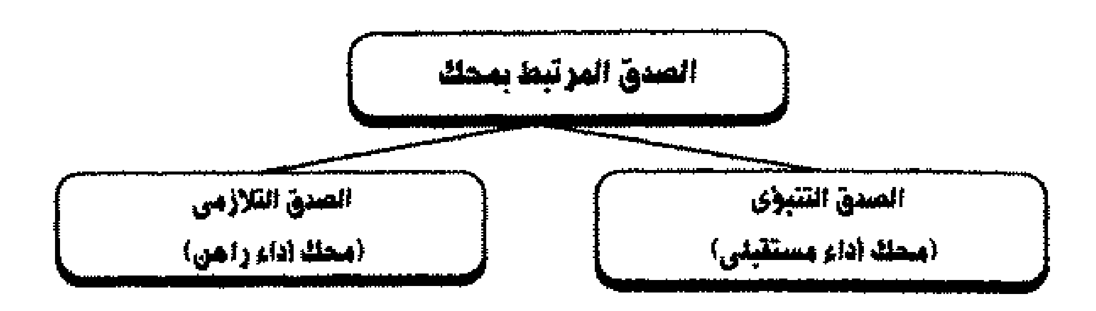
فإلى جانب الأحكام التقييمية للمقياس المحك ينبغي أيضًا التحقق من ثبات درجاته وخلوها من التحيز ومواءمة المقياس للأداء المطلوب. فإذا اعتبرنا تقديرات المشرفين على العمل محكًا لأداء العاملين، فإن هذا يتطلب التحقق من دقة هذه التقديرات وعدم تحيزها، ومدى مواءمتها لطبيعة العمل وخصائص العاملين؛ لكي تكون ذات فائدة في الحكم على صدق القرار الذي يتخذ في ضوء درجات الاختبار المنبئ.

أنواع الصدق المرتبط بمحك:

تميز أدبيات القياس والتقويم بين نوعين من الصدق المرتبط بمحك أحدهما يسمى الصدق التنبؤي Predictive Validity والآخر يسمى الصدق التلازمي Concurrent Validity. ويعتمد هذا التمييز على إمكانية توافر محكات أداء لاحقة أو مستقبلية، ومحكات أداء راهنة.

وترجع أهمية هذا التمييز إلى أن المدة الزمنية التي تفصل بين تطبيق الاختبار التنبؤي والاختبار أو المقياس المحك تؤثر في درجة العلاقة بين الاختبارين. فالاختبار الذي يمكن باستخدامه تقدير الكفاية الحالية لدى الطلاب أو العاملين ربما لا نستطيع باستخدامه التنبؤ بالأداء المستقبلي أو اللاحق والعكس أيضًا يكون صحيحًا.

والشكل التخطيطي (٥-٢) التالي يوضح هذين النوعين من الصدق:



شكل (٢-٥) يوضح نوعى الصدق المرتبط بمحك

الصدق التنبؤى :

تتعلق أدلة الصدق التنبؤى Predictive Validity بتقدير مدى صلاحية الاختبار فى التنبؤ بأداء الفرد المستقبلى الذى يقاس باختبار محك باستخدام درجات اختبار يطبق عليه فى الوقت الحاضر، وبعبارة أخرى تتعلق هذه الأدلة بمدى اقتران تباين درجات اختبار منبئ بتباين درجات اختبار محك، أى درجة العلاقة بين درجات اختبار ونمط معين من السلوك المستقبلى مما يجعل من الممكن التنبؤ بهذا السلوك. وهذا يتطلب مرور مدة زمنية بين الحصول على الدرجات فى الاختبار التنبؤى ودرجات أو تقديرات الاختبار المحك الذى نقدِّر فى ضوئه صدق القرار.

فإذا وجدنا ارتباطًا مرتفعًا بين اختبار الاستعدادات الفنية الذى يطبق على الطلاب المتقدمين للقبول بكليات الفنون الجميلة والمعدل العام لتقديراتهم فى نهاية السنة الأولى بهذه الكليات، فإن هذا يمكن أن يقدم دليلاً على الصدق التنبؤى للقرارات المتعلقة بقبول الطلاب للالتحاق بهذه الكليات أو رفضهم.

فالمحك فى هذه الحالة هو المعدل العام للتقديرات كمؤشر لنجاح الطلاب. وربما لا يتفق الجميع على هذا المحك، فاختبار الاستعدادات الفنية ربما يرتبط ارتباطًا مختلفًا بمحكات مختلفة.

ولذلك يمكن أن يكون هناك معاملات صدق تنبؤى مختلفة لاختبار واحد، أى أنه لا يوجد معامل صدق تنبؤى وحيد يمثل صدق مجموعة من المقاييس أو البيانات المستمدة منها.

طرق تقدير الصدق التنبؤى :

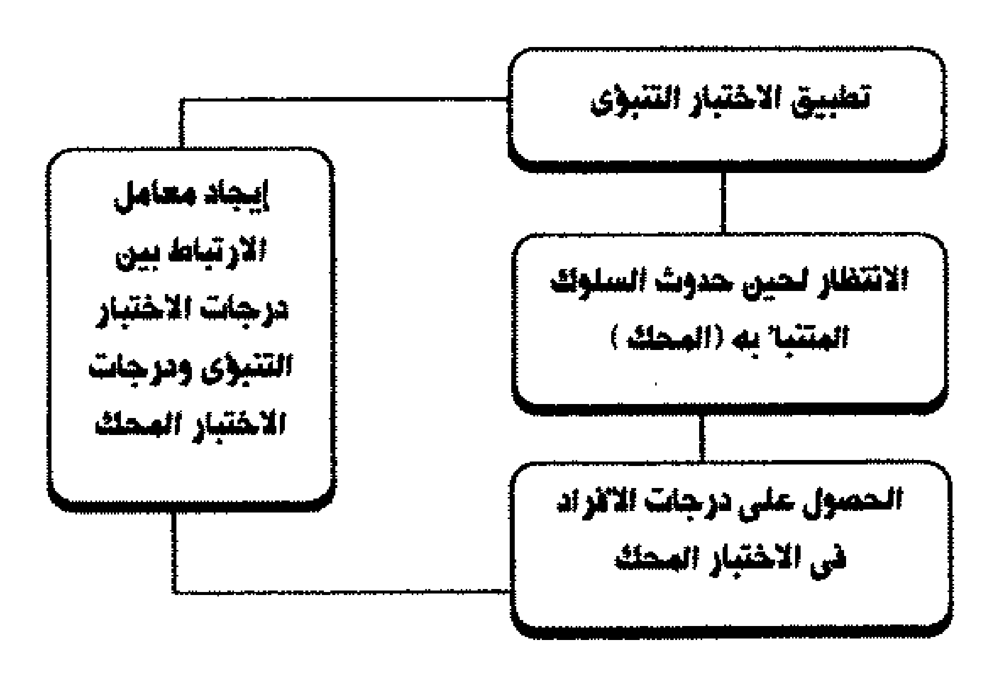
تتعدد طرق تقدير الصدق التنبؤى اعتماداً على مجالات استخدام الاختبارات والمقاييس ونوع القرارات التى تسترشد بالبيانات المستمدة من دراسات الصدق، غير أننا سوف نناقش بإيجاز بعض الطرق شائعة الاستخدام فى هذا الشأن.

طريقة تعتمد على الارتباط بين الاختبار التنبؤى واختبار المحك:

يمكن تقدير الصدق التنبؤى فى المجالات التربوية ومجالات الأعمال والوظائف عن طريق متابعة تطبيق اختبار معين على جميع المتقدمين فى شهور مختلفة دون أن يستخدم الاختبار فى أغراض الانتقاء وإنما يستخدم فى أغراض البحث، بل وأحياناً لا يجرى تصحيح الاختبار إلا بعد إلحاق بعض هؤلاء المتقدمين بالعمل.

فالذين تم إلحاقهم يستمرون فى العمل مدة زمنية معينة ثم تجمع بيانات باستخدام الاختبار أو المقياس المحك، أو أى أداة لجمع معلومات مناسبة عن هؤلاء المقبولين، وتحسب قيمة معامل الارتباط بين درجات كل من الاختبار التنبؤى والاختبار أو المقياس المحك.

ويمكن توضيح خطوات هذه الطريقة بالشكل التخطيطى (٥ ـ ٣) التالى:



شكل (٥ ـ ٣) يوضح خطوات تقدير الصدق التنبؤى للاختبار

فإذا كانت هذه القيمة مرتفعة بدرجة كافية يمكن اعتبار قرارات القبول أو الرفض صادقة، ويصبح الاختبار التنبؤى جزءًا من عملية الانتقاء لشغل وظائف أو أعمال معينة.

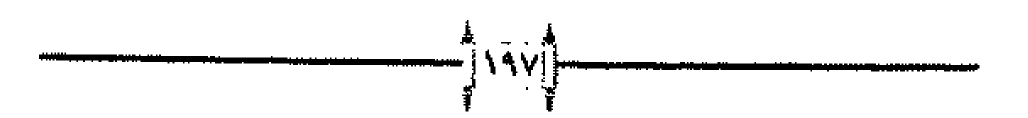
وهنا يفترض أن درجة صلاحية الاختبار فى الواقع الميدانى لا تختلف عن درجة صلاحيته عندما تم تقدير الصدق التنبؤى لأغراض البحث. غير أن هذا الافتراض يصبح غير صحيح إذا حدث تغير ملحوظ فى الظروف المحيطة بالعمل، حيث يصبح الصدق التنبؤى عندئذ موضع تساؤل.

ويطلق على معامل الارتباط بين درجات الاختبار التنبؤى ودرجات محك الأداء أو محك النجاح فى وقت لاحق معامل الصدق التنبؤى. وينبغى أن تشتمل اختبارات المحك أو أى أدوات أخرى على عينة ممثلة للمهارات أو الصفات المطلوبة للعمل بحيث يتعلق محتوى الاختبار أو الأداة المحكة بالسلوك المتوقع.

وينبغى مراعاة أن العوامل التى تؤثر فى معامل الارتباط والتى ناقشناها فى الفصل الثالث تؤثر أيضًا فى معامل الصدق التنبؤى، كما أنه يمكن تفسير قيم معامل الصدق التنبؤى بنفس الطريقة التى فسرنا بها قيم معامل الارتباط فى الفصل نفسه، وذلك باستخدام مفهوم التباين المشترك لدرجات كل من الاختبار التنبؤى واختبار المحك الذى نحصل عليه بتربيع قيم معامل الصدق التنبؤى. فإذا كانت قيمة هذا المعامل ٠,٧١ مثلا فإن هذا يعنى أن ٥٠٪ من تباين درجات الاختبار المحك (أى نصف مقدار التباين) يعزى إلى تباين درجات الاختبار التنبؤى.

وينبغى الإشارة هنا إلى أنه لا ينبغى أن نتوقع الحصول على قيم مرتفعة لمعامل الارتباط بين درجات كل من الاختبارين التنبؤى والمحك فى دراسات الصدق التنبؤى، وذلك لتعقد وتشابك السلوك الإنسانى مما يجعل التنبؤ الدقيق بأداء الأفراد فى مختلف المواقف التربوية والعملية باستخدام اختبار تنبؤى واحد أو أكثر أمرًا غير واقعى.

ولذلك يرى نانلى **Nunnally** أنه من الأفضل تفسير قيم معامل الصدق فى ضوء التحسن المحتمل فى متوسط أداء الأفراد الذى يمكن الحصول عليه باستخدام الاختبار التنبؤى، فالاختبار التنبؤى الذى يرتبط بالمحك ارتباطًا متوسطًا نسبيًا مثل ٠,٤٠ يمكن أن يفيد فى التنبؤ بهذا التحسن المحتمل فى بعض المواقف التربوية والعملية، فالتنبؤ بأداء فرد معين يشوبه قدر من الخطأ أكبر من التنبؤ بمتوسط أداء مجموعة من الأفراد.



طريقة تعتمد على استخدام معادلة الانحدار فى التنبؤ بدرجات الاختبار المحك:

سبق أن أشرنا في الفصل الثالث إلى أنه يمكننا التنبؤ بقيمة أحد المتغيرات إذا علمنا قيمة متغير آخر باستخدام صيغة معادلة خط الانحدار Regression Equation التى تعتمد على قيمة معامل الارتباط بين المتغيرين.

لذلك يمكن استخدام هذه المعادلة فى التنبؤ بدرجة الفرد فى الاختبار المحك بمعلومية درجته في الاختبار التنبؤى اعتمادًا على قيمة معامل الصدق التنبؤى. غير أن هذا يتطلب أن تكون العلاقة بين درجات الاختبارين مستقيمة. ويمكن تقدير أخطاء التنبؤ باستخدام صيغة الخطأ المعيارى للتنبؤ (٣ ـ ١٢) التى أشرنا إليها أيضًا فى الفصل نفسه.

ويمكن التوصل إلى معادلة خط الانحدار فى مرحلة دراسة صدق الاختبار المستخدم فى التنبؤ، إذ يمكن استخدام قيمة معامل الارتباط بين هذا الاختبار والاختبار المحك، وكذلك قيمة المتوسط والانحراف المعيارى لدرجات كل منهما فى التوصل إلى معادلة الانحدار

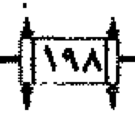
$$\hat{ص} = \bar{ص} + ر \times \frac{ع_{ص}}{ع_{س}} (س - \bar{س}).$$

وتشتمل هذه المعادلة على متغيرين أحدهما المتغير المتنبئ (س) والآخر المتغير المتنبأ به ($\hat{ص}$)، أما بقية رموز المعادلة وهى ($\bar{س}$)، ($ع_{س}$)، ($\bar{ص}$)، ($ع_{ص}$)، ر فإنه يتم حساب قيمة كل منها فى مرحلة الدراسة الإمبيريقية للصدق.

ويلاحظ أن ($\bar{ص}$) ترمز لمتوسط درجات عينة الأفراد المستخدمة فى دراسة الصدق، ($\bar{س}$) ترمز لمتوسط درجات اختبار المحك المتنبأ بها من هذه العينة التى تكون درجات أفرادها فى الاختبار التنبؤى (س).

ونظرًا لأننا نتوقع عند استخدام الاختبار التنبؤى فيما بعد على عينات أخرى مماثلة أن تختلف درجاتهم المتنبأ بها ($\hat{ص}$) عن الدرجات المتنبأ بها لأفراد عينة دراسة الصدق، لذلك فإنه يمكن تقدير هذه الأخطاء باستخدام صيغة الخطأ المعيارى للتنبؤ $ع_{ص}\sqrt{1 - ر^{2}}$. وقد سبق أن أوضحنا كيفية تفسير القيمة التقديرية لهذا الخطأ فى الفصل الثالث.

ويمكن استخدام معادلة خط الانحدار بعد ذلك فى التنبؤ بدرجات أفراد عينات أخرى مماثلة للعينة التى استمدت منها بيانات هذه المعادلة فى الاختبار المحك، بمعلومية درجاتهم فى الاختبار التنبؤى.



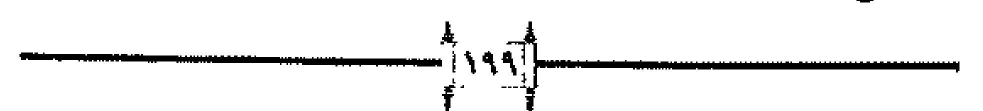
وكما سبق أن ذكرنا فإننا نتوقع أن تكون القيم التقديرية للخطأ المعيارى للتنبؤ كبيرة وذلك لاعتمادنا على اختبار تنبؤى واحد. ولذلك ربما يفضل استخدام أكثر من اختبار تنبؤى واحد، وضم درجاتها معًا بأوزان مختلفة تتناسب وأهميتها فى التنبؤ بدرجات اختبار المحك.

ولإجراء ذلك نوجد معامل الارتباط المتعدد Multiple Correlation بين التركيب الخطى لدرجات الاختبارات التنبؤية ودرجات اختبار المحك، أو استخدام أسلوب الانحدار الخطى المتعدد Multiple Linear Regression فى تحديد هذه الأوزان. وقد وضح المؤلف هذه الأساليب الإحصائية وكيفية استخدامها فى مرجع آخر.

غير أنه ينبغى فى هذه الحالة مراعاة عدم استخدام عدد كبير من الاختبارات التنبؤية، والاقتصار على اختبارين أو ثلاثة، نظرًا لأن كثرة الاختبارات المرتبطة بالاختبارالمحك لا تسهم إسهامًا ملحوظًا فى تفسير تباين درجات هذا الاختبار أكثر مما تسهم به ثلاثة اختبارات مثلاً. ولكن ينبغى مراعاة أن يكون الارتباط بين الاختبارات التنبؤية منخفضًا، وارتباطها باختبار المحك مرتفعًا لكى تسهم فى التنبؤ إسهامًا جيدًا، وبذلك تكون أكثر جدوى فى التنبؤ ولا تتطلب وقتًا أو جهدًا كبيرا فى التطبيق.

وعلى الرغم من أن الاختبارات التنبؤية تشتمل على مفردات يُفترض أنها تُميز بين المجموعات أو تتنبأ بالنواتج، إلا أن المفردات المنتقاة التى تميز بين أفراد مجموعة معينة ربما لا تميز بين أفراد مجموعة أخرى، فالارتباط بين درجات اختبار تنبؤى معين ودرجات اختبار محك فى المجموعة الأصلية التى استخدمت فى بناء الاختبار التنبؤى لا تؤدى إلى معاملات صدق مرتبط بمحك معقولة ولكن تتذبذب من مجموعة إلى أخرى. لذلك يُفضل إجراء دراسة صدق مستعرض Cross Validation على عينة جديدة مستقلة من الأفراد تكون مستمدة من المجتمع الذى سيطبق عليه الاختبار فيما بعد. وهذا الأسلوب يهدف لإعادة فحص بيانات الصدق على المجموعة الجديدة، فالمفردات التى كانت مميزة نتيجة للصدفة سوف نجد أنها أصبحت غير مُميِّزة عند إعادة تطبيق الاختبار.

أما إذا بينت دراسات الصدق أن النتائج متماثلة فى المجموعتين، فإن هذا يعد دليلا على العلاقة بين الاختبار المنبئ والمحك، ولا ترجع إلى عوامل الصدفة. ولكن إذا كانت النتائج مختلفة فإنه ينبغى بناء مفردات جديدة ترتبط بالمحك.



بعض المشكلات المتعلقة بالمحك :

تعتمد دقة تقدير الصدق المرتبط بمحك على نوع المحك الذى يستخدم فى مجالات معينة. واختيار المحك المناسب والاختبار أو المقياس الذى يزودنا بدرجات أو تقديرات أو معلومات عن سلوك الفرد أو أدائه فى موقف معين يعد من المشكلات الأساسية التى تواجه القائم بإعداد الاختبارات التنبؤية والمسئولين عن اتخاذ قرارات الانتقاء والتصنيف والتسكين وربما التشخيص أيضًا. فالمحكات تختلف باختلاف المجال موضع الاهتمام، إذ ربما تستخدم درجات الاختبارات التحصيلية التى تطبق فى نهاية مدة دراسية معينة، أو تقديرات المعلمين حول سلوكيات تلاميذهم وإنجازاتهم، أو المعدل العام كمحكات لتقييم الصدق التنبؤى للاختبارات فى مجالات انتقاء الطلاب أو تسكينهم فى برامج تعليمية مناسبة. وربما تستخدم تقديرات المشرفين على العمل، والتقييم الذاتى، ومؤشرات الإنتاج وسجل الخدمة، وتقديرات الدورات التدريبية فى مجالات انتقاء العاملين أو إلحاقهم بوظائف أو مهن أو أقسام مناسبة، أو ترقيتهم إلى وظائف أو مهن عليا.

واستخدام محك مناسب يتطلب توافر بعض الشروط الأساسية، من أهمها كما سبق أن ذكرنا مواءمته لنوع القرار المراد اتخاذه ولطبيعة السلوك المراد التنبؤ به. فالسلوك الذى يتضمنه المحك مثل تقديرات المعلمين فيما يتعلق بتحصيل تلاميذهم لا تكون مقياسًا موائمًا لنجاح التلاميذ إلا إذا عكست هذه التقديرات ما يتضمنه النجاح من خصائص معينة، إذ ربما يختلف مفهوم النجاح من زمان إلى آخر، فالطالب الذى يعد ناجحًا الآن ربما لا يكون كذلك مستقبلاً، وهذا يعتمد على ما نقصده بمفهوم النجاح وما يتضمنه من خصائص.

فعندما يستند تقييم النجاح في هذه الحالة على تقديرات المعلمين، فإن هذا المحك سوف يكون متحيزًا. وللتغلب على مشكلة تحيز المحكات وبخاصة التى تعتمد على الأحكام القيمية والتقديرات ينبغى تحديد العناصر والمكونات الأساسية التى ينبغى أن تسترشد بها هذه الأحكام تحديدًا إجرائيًا للتقليل من الذاتية وتأثير الهالة Halo Effect أى الأفكار المسبقة عن الفرد أو عن درجته في الاختبار التنبؤى، وجميع عوامل التحيز قدر الإمكان.

ومن المشكلات الأساسية أيضًا المتعلقة بالمحكات ثبات درجات المحك، فالاختبار أو المقياس الذى يستخدم فى تقدير الصدق التنبؤى يُفترض أن درجاته تتسم بالثبات أو الاتساق، غير أن واقع الأمر ليس كذلك. فكثير من مقاييس المحك تعتمد

على التقارير الذاتية، والتقديرات، والمقابلات الشخصية، وجميعها تعتمد على الأحكام الشخصية بدرجات متفاوتة، وهذه المقاييس تؤثر فيها كما ذكرنا عوامل تحيز متعددة مما يقلل من ثبات درجاتها.

ويمكن استخدام بعض الأساليب الإحصائية التى تساعد فى تصحيح الأخطاء التى ربما تكون قد أدت إلى خفض معامل ثبات درجات كل من الاختبار التنبؤى واختبار المحك. ومن بين هذه الأساليب الصيغة الإحصائية التالية:

أقصى معامل ارتباط بين الاختبار التنبؤى واختبار المحك =

$$\frac{\text{معامل الصدق التنبؤى}}{\text{معامل ثبات الاختبار التنبؤى} \times \text{معامل ثبات الاختبار المحك}} \quad \ldots (٥ - ١)$$

فإذا كان معامل الصدق التنبؤى للاختبار التنبؤى ٠,٣٠ مثلا، ومعامل ثبات درجاته ٠,٨٥، ومعامل ثبات درجات المحك ٠,٥٠، فإن أقصى قيمة لمعامل الارتباط بين درجات كل من الاختبارين =

$$\frac{٠,٣٠}{٠,٥٠ \times ٠,٨٥} = ٠,٧١$$

وهذه القيمة تساوى قيمة معامل الارتباط بين درجات الاختبار التنبؤى ودرجات اختبار المحك إذا افترضنا أن درجات كل منهما غير مشوبة بأى نوع من الخطأ، أى أن ثبات درجات كل منهما تام. ولكن ينبغى أن يميز دليل الاختبار بين القيمة الأصلية لمعامل الصدق وقيمته المتوقعة بعد إجراء عملية التصحيح.

والصيغة (٥ ـ ١) تبين القيمة المتوقعة لمعامل الارتباط بين درجات الاختبارين إذا ارتفعت قيم معامل ثبات درجات الاختبار التنبؤى والاختبار المحك. فإذا لم نجد ارتفاعًا ملحوظًا فى القيمة المتوقعة عما كانت عليه بالفعل، فإنه ينبغى إعادة النظر فى الاختبار التنبؤى أو فى طريقة بنائه، نظرًا لأنه فى هذه الحالة لا يُعزى إليه نسبة كبيرة من تباين الدرجات الحقيقية للاختبار المحك.



طريقة تعتمد على جداول التوقع :

تستخدم هذه الطريقة في وصف العلاقة بين درجات اختبار تنبؤي ودرجات اختبار المحك. وتعتمد على جداول يمكن تكوينها تسمى جداول التوقع Expectancy Tables ، وتنظم بشكل يسمح بتقدير أرجحية حصول فرد على درجة أو تحقيقه لمستوى معين من الكفاية في اختبار محك إذا علمنا درجته في اختبار تنبؤي. أي أن هذه الجداول تفيد في تقدير احتمالات نجاح الأفراد الذين حصلوا على درجات معينة في الاختبار التنبؤي. فمشكلة انتقاء الأفراد تتعلق بالتنبؤ بمستوى معين من مستويات الكفاية أو النجاح في دراسة أو مهنة معينة، أي أن التنبؤ هنا يكون تصنيفيًا. فالمعلم أو المسئول عن العمل لا يستطيع على وجه التحديد معرفة ما إذا كان الطالب أو العامل سيؤدي أداء جيدًا في دراسته أو مهنته، ولكن البيانات المتعلقة بالصدق التنبؤي يمكن أن تساعده في تقدير أرجحية ذلك استرشادًا بالنسب المئوية أو التكرارات النسبية لعدد الحالات التي حققت مستويات نجاح معينة والمدونة في جداول التوقُّع الناتجة عن دراسات الصدق.

ولتكوين جدول التوقُّع ينبغي الحصول على توزيع درجات كل من الاختبار التنبؤي واختبار المحك كما سبق أن أوضحنا، وتقسيم درجات كل من الاختبارين إلى عدد من الأقسام . ويفضل أن تكون هذه الأقسام متساوية في عددها، وليس في مدى درجاتها.

ونوجد عدد الأفراد الذين ينتمون إلى كل قسم من هذه الأقسام، وكذلك التكرارات النسبية أو النسب المئوية من الأفراد في كل قسم. ثم نكون جدول التوقع الذي يوضح العلاقة الملاحظة بين النسب المئوية من الأفراد في كل قسم من أقسام اختبار المحك لكل فئة من فئات درجات الاختبار التنبؤي.

ويمكن إجراء ذلك بتمثيل فئات درجات الاختبار التنبؤي على المحور الأفقي ، وأقسام درجات اختبار المحك على المحور الرأسي، وتدوين النسب المئوية من الأفراد الذين ينتمون إلى فئة معينة من فئات درجات الاختبار التنبؤي وينتمون في الوقت نفسه إلى قسم معين من أقسام درجات اختبار المحك. ويمكن توضيح ذلك بالمثال الافتراضي التالي:

نفترض أن الاختبار التنبؤي هو اختبار القبول بالكليات واختبار المحك هو المعدل التراكمي للطالب في نهاية السنة الأولى الجامعية، ونفترض أنه أجريت دراسة الصدق التنبؤي للبيانات المستمدة من هذين الاختبارين، فعندئذ يمكن تكوين جدول

التوقع وفقا للخطوات السابقة، وذلك بتمثيل فئات درجات اختبار القبول على المحور الأفقي ، وأقسام أو فئات المعدل التراكمي على المحور الرأسي، وتكوين جدول توزيع تكراري لعدد الطلاب في كل خلية من خلايا الجدول (٥ ـ ١) التالي:

المجموع	درجات اختبار المحك (المعدل التراكمي)					درجات الاختبار التنبؤي (اختبار القبول)
المجموع	٤,٠	٣,٠	٢,٠	١,٠	صفر	
٨	١	٥	٢			٦٠-٦٤
٩	١	٦	١	١		٥٥-٥٩
١٦	٢	٧	٥	٢		٥٠-٥٤
١٥	٢	٤	٤	٥		٤٥-٤٩
١٦		٣	٧	٤	٢	٤٠-٤٤
١٦		١	٦	٧	٢	٣٥-٣٩
٩		٣	٣	٢	١	٣٠-٣٤
٧			١	٤	٢	٢٥-٢٩
٤				٣	١	٢٠-٢٤
١٠٠	٦	٢٩	٢٩	٢٨	٨	المجموع

جدول (٥ ـ ١) جدول توزيع تكراري للمعدل التراكمي لكل فئة من فئات درجات اختبار القبول

والخطوة التالية : نحول عدد الطلاب في كل خلية إلى نسب مئوية من العدد الكلي لطلاب كل فئة من فئات درجات اختبار القبول كما في جدول (٥ ـ ٢) التالي:

درجات الاختبار التنبؤى (اختبار القبول)	درجات اختبار المحك (المعدل التراكمى)					
	صفر	١,٠	٢,٠	٣,٠	٤,٠	المجموع
٦٤-٦٠			٢٥	٦٣	١٢	١٠٠
٥٩-٥٥		١١	١١	٦٧	١١	١٠٠
٥٤-٥٠		١٣	٣١	٤٤	١٢	١٠٠
٤٩-٤٥		٣٣	٢٧	٢٧	١٣	١٠٠
٤٤-٤٠	١٢	٢٥	٤٤	١٩		١٠٠
٣٩-٣٥	١٢	٤٤	٣٨	٦		١٠٠
٣٤-٣٠	١٢	٢٢	٣٣	٣٣		١٠٠
٢٩-٢٥	٢٩	٥٧	١٤			١٠٠
٢٤-٢٠	٢٥	٧٥				١٠٠

جدول (٥ ـ ٢) جدول توقع للمعدل التراكمى لكل فئة من فئات

درجات اختبار القبول

وبالنظر إلى جدول (٥ ـ ١) يتضح أن ١٦ طالبًا حصلوا على درجات فى اختبار القبول تتراوح بين ٤٠، ٤٤، والمعدل التراكمى لطالبين منهم (صفر)، ولأربعة طلاب (١,٠) ولسبعة طلاب (٢,٠) ولثلاثة منهم (٣,٠) .

ويوضح جدول (٥ ـ ٢) النسب المئوية لعدد طلاب كل فئة من فئات درجات اختبار القبول المناظرة لمستويات المعدل التراكمى. وإذا نظرنا إلى الخلايا المناظرة للأعداد السابقة نجد أنها تناظر النسب المئوية ١٢ ، ٢٥ ، ٤٤ ، ١٩ على الترتيب.

ويمكن تفسير هذه النسب على أنها احتمالات حصول طلاب على معدل تراكمى معين إذا علمنا درجاتهم فى اختبار القبول.

لذلك يمكننا القول بأن ١٢ ٪ من الطلاب الذين تتراوح درجاتهم فى اختبار القبول بالكليات بين ٤٠ ، ٤٤ يتوقع أن يحصلوا فى نهاية السنة الجامعية الأولى على معدل تراكمى (صفر) ، ٢٥ ٪ منهم يتوقع أن يحصلوا على معدل تراكمى (١)، وهكذا.

وإذا اعتبرنا الطلاب الناجحين هم الذين يزيد معدلهم التراكمي عن (٢,٠)، بينما يُعد راسبًا من يقل معدله التراكمي عن (٢,٠)، فبالنظر إلى جدول (٥ ـ ١) يلاحظ أن: ٢٨ + ٧ = ٣٥ طالبًا من بين المجموع الكلي للطلاب (١٠٠ طالب) يُعدون ناجحين، فإذا تم قبول العدد الكلي للطلاب فإنه يُتوقع أن ينجح منهم ٣٥ طالبًا، أي ٣٥٪، أما النسبة الباقية وهي ٦٥٪ فيُتوقع رسوبهم. ويوضح ذلك الخط الرأسي الفاصل عند المعدل التراكمي (٢,٠) في الجدول.

وإذا تم قبول جميع الطلاب الذين حصلوا على الدرجة ٤٥ وما فوقها في اختبار القبول، فإن عدد المقبولين يكون ٢٨ + ٢٠ = ٤٨ طالبًا، ويُتوقع أن ينجح منهم ٢٨ طالبًا. ويوضح ذلك المربعين العلويين الأيمن والأيسر، ونسبة هؤلاء الطلاب الناجحين إلى عدد المقبولين تسمى نسبة النجاح Success Ratio، أي $\frac{٢٨}{٤٨}$ = ٥٨ ٪ استنادًا إلى درجتي القطع المختارتين، أي ٤٥، ٢,٠.

وإذا قسمنا عدد الطلاب المقبولين الذين تفوق درجاتهم الدرجة ٤٥ في اختبار القبول (٤٨) على العدد الكلي للطلاب (١٠٠)، أي ٤٨٪ ، فإن هذه النسبة تسمى نسبة القبول Success Ratio. وينبغي ملاحظة أنه كلما زاد عدد المقبولين فإن نسبة القبول تنخفض، كما ينبغي ملاحظة أنه إذا كان معامل الصدق تامًا أي مساويًا الواحد الصحيح فإن جميع القيم في خلايا الجدول تتراكم بحيث تقع على خط مستقيم، وعندئذ يمكن التنبؤ بدرجة المحك بمعلومية درجة اختبار القبول تنبؤًا تامًا بغض النظر عن موقع درجة القطع، وبذلك تكون نسبة النجاح واحدًا صحيحًا، أي لا يحتمل رسوب أي طالب تم قبوله، أما من رفض قبوله فإننا نتوقع أنه كان سيرسب إذا تم قبوله عن طريق الخطأ.

أي أن هناك علاقة تفاعلية بين صدق التنبؤ ونسبة النجاح ونسبة القبول. وقد وضح هذه العلاقة تايلر وراسل Tayler - Russell في جداول تتضمن قيمًا مختلفة لكل من الصدق ونسبة النجاح ونسبة القبول. ويمكن باستخدام هذه الجداول تقدير نسبة الأفراد الناجحين الذين يقبلون استنادًا إلى اختبارات تنبؤية متباينة في درجة صدقها، ونسب قبول مختلفة.

ونظرًا لأن كلا من نسبة النجاح ونسبة القبول تعتمد على درجة القطع في الاختبار التنبؤي ، فإنه ينبغي مراعاة التصنيف الخاطئ الذي يترتب على هذه الدرجة عند تقييم فاعلية اختبار كأداة فحص. ونعني بالتصنيف الخاطئ رفض قبول فرد كان من الممكن نجاحه إذا التحق بالدراسة أو بالعمل ويسمى الخطأ الموجب False Positive، وكذلك نسب النجاح والرسوب لمجموعة الأفراد.

وتختلف هذه النسب باختلاف طبيعة الدراسة أو العمل الذى سيوكل إلى الأفراد المقبولين. فبعض أنواع العمل تتطلب كفايات مرتفعة وبالتالى تكون درجة القطع مرتفعة ارتفاعًا نسبيًا مما يؤدى إلى الخطأ الموجب، وهنا يمكن أن نقلل من هذا الخطأ بخفض درجة القطع. ولكن هناك أعمال ربما تستدعى قبول عدد كبير من الأفراد على الرغم من احتمال فشل بعضهم وهو ما يسمى بالخطأ السالب **False Negative**.

وعمومًا، فإن موقع درجة القطع يتحدد بناء على حاجة العمل من الأفراد العاملين، وعدد الأماكن الشاغرة ومدى الحاجة الملحة لشغلها.

ويمكن للمرشدين التربويين فى المدرسة الثانوية تكوين جداول توقُّع مثل جدولى (٥ ـ ١) ، (٥ ـ ٢)، لمعاونة الطلاب فى اتخاذ قرارات تتعلق بالكليات التى يزمعون الالتحاق بها مع مراعاة الأخطاء التى تنجم عن التنبؤ باستخدام هذه الجداول كما هو الحال فى أخطاء الانحدار.

ولتفسير درجة أحد الطلاب نحدد موقعها فى العمود الأول للفئات فى جدول التوقُّع (٥ ـ ١) ، ونقرأ النسبة المئوية المدونة فى الخلية الناتجة من تقاطع الصف الذى تقع فيه درجته فى اختبار القبول مع العمود الذى يناظر المعدل التراكمى المتوقع، وذلك لتقدير احتمال حصوله على هذا المعدل فى نهاية السنة الجامعية الأولى.

ومن الجدير بالذكر أنه يمكن تكوين جداول توقُّع لاختبارين من اختبارات التنبؤ بدلا من اختبار واحد، فمثلاً يمكن استخدام اختبار القبول ودرجات نهاية العام فى الثانوية العامة للتنبؤ بالمعدل التراكمى للطالب فى نهاية السنة الجامعية الأولى، واتباع الخطوات السابقة فى تكوين جدول التوقُّع.

غير أنه فى هذه الحالة ينبغى تحديد درجة قطع لكل اختبار على حدة بحيث يمكن استبعاد الأفراد من ذوى الدرجات المنخفضة فى أى اختبار؛ نظرًا لأن الدرجة المركبة من درجات الاختبارات تعوّض بعضها البعض الآخر.

والخلاصة أن جداول التوقع تتضمن بيانات تتعلق بالصدق التنبؤى، ويستطيع المسئولون عن انتقاء الأفراد استخدامها فى تفسير درجات الاختبارات التنبؤية بعبارات احتمالية فى ضوء نواتج معينة متوقعة.

ولكن يعاب على هذه الجداول استخدام فئات درجات الاختبار التنبؤى لكل قسم من أقسام درجات المحك مما يقلل من الدقة فى التفسير، كما أن بيانات هذه الجداول تُستمد من عينات من الأفراد فى الماضى، وربما تختلف خبراتهم عن خبرات الأفراد الذين نتنبأ بدرجاتهم فى المحك بعد ذلك وهذا يتطلب الحيطة والنظرة الفاحصة عند

تفسير البيانات المدونة فى هذه الجداول ؛ نظراً لأن الاختبارات التنبؤية لا تقيس خصائص مهمة أخرى لدى الأفراد مثل المثابرة ، والمبادأة، والدافعية، مما قد يؤثر فى نتائج التوقع.

كذلك تُستمد هذه الجداول عادة من عينة عدد أفرادها قليل نسبيًا مما يجعل هناك خلايا فارغة أو قليلة العدد كما فى الجدول (٥ ـ ١)، مما يؤدى إلى عدم اتساق قيم الاحتمالات.

ونظراً لأن هذه الجداول استُمدت من مجموعة من الأفراد، فإن تفسير هذه القيم ينطبق على المجموعة ككل. ففى الجدول السابق وجدنا أن ١٩ ٪ من طلاب المجموعة الذين تتراوح درجاتهم بين ٤٠ ، ٤٤ يُتوقع حصولهم على معدل تراكمى (٣)، ومعنى ذلك أنه إذا تم قبول الطلاب الذين تتراوح درجاتهم فى اختبار القبول بين ٤٠ ، ٤٤ فإننا نتوقع أن ١٩ ٪ من هؤلاء الطلاب سوف ينجحون فى دراستهم بالكلية، فإذا حصل أحد الطلاب على درجة فى اختبار القبول تتراوح أيضًا بين ٤٠، ٤٤ فهو إما سيكون ناجحًا أو راسبًا، ويصبح احتمال نجاحه ١٩٪ غير ذى معنى. فالاحتمال ١٩٪ يشير إلى توقع النجاح لمجموعة الأفراد الذين حصلوا على درجات تنحصر بين درجتين معينتين فى اختبار القبول، أى أن التوقع أو التنبؤ هنا لا يتعلق بكل فرد على حدة، وإنما يتعلق بمتوسط المجموعة؛ إذ ربما لا يؤدى الأفراد أداء يتفق وتوقعاتنا، وإنما ربما نجد بعضهم يؤدى أداء أفضل مما توقعنا، لذلك لا يجب أن نخبرهم بتوقعنا عن أدائهم المنخفض.

ومع هذا فإن استخدام جداول التوقع يسهم فى ترشيد قرارات الانتقاء بدرجة أفضل مما لو تم الانتقاء دون استخدام أساليب تنبؤ علمية.

طريقة تعتمد على نظرية المنفعة :

اتضح لنا أن جداول التوقُّع تفيد فى ترشيد قرارات انتقاء الأفراد، كما اتضح كيف تتأثر هذه القرارات بدرجة الصدق التنبؤى لأدوات القياس المستخدمة وبنسبة القبول ونسبة النجاح، غير أن هذه القرارات تتأثر أيضًا بالمنفعة النسبية Utility للنواتج المتوقعة وما يتعلق بهذه النواتج من كلفة وعائد. فنظرية القرارات تستند إلى نظام لتعيين قيم لمختلف النواتج من حيث المنفعة النسبية، وهذا يتطلب تحليلاً منظمًا لعلاقة النواتج بكلفتها وعائدها، وهذا التحليل يعد ذى فائدة كبيرة فى مجال انتقاء الأفراد باستخدام الاختبارات والمقاييس التنبؤية.

فكثير من المؤسسات تُعنى بانتقاء الأفراد من خلال برامج تختص بعمليات الفحص والانتقاء، وتتطلب هذه البرامج تمويلاً لإعدادها وتطويرها تطويرا مستمرا. وعندما يقبل فرد ويلتحق بالعمل فإنه يؤدى منفعة للمؤسسة إذا كان ناجحا مما يعد بمثابة عائد للمؤسسة، أما إذا لم يتبين نجاحه فإنه بذلك يعد بمثابة فاقد للمؤسسة. وفي جميع هذه الأحوال تبدو أهمية تحليل كلفة وعائد القرار الذى يتخذ وما يترتب عليه.

وهذا يتطلب تقدير قيم المنفعة لكل من هذه القرارات متضمنة الكلفة المالية والعائد. فمثلاً عائد العامل الناجح على المؤسسة ربما يتم تقديره بطرح المرتب الذى يتقاضاه وكلفة الأجهزة أو الآلات التى يستخدمها فى العمل من جملة عائد إنتاجه. وربما تُدخل أيضًا بعض المتغيرات الأخرى مثل تأثيره فى إنتاجية زملائه، أو مدة انقطاعه عن العمل بسبب ظروف طارئة، وهكذا. أما كلفة العامل الذى يتبين عدم فاعليته فإنه يمكن تقديرها من كلفة برنامج الانتقاء، وكلفة إلحاق فرد آخر للقيام بعمله، وفاقد الإنتاجية التى ترتبت على ذلك.

والخلاصة أن تقدير الصدق التنبؤى لقرارات الانتقاء ينبغى أن يأخذ بعين الاعتبار كلفة وعائد البرنامج الاختبارى وما يترتب عليه من منفعة للمؤسسة، ومحاولة ترجمة ذلك اقتصاديًا، مما يجعل مفهوم الصدق التنبؤى له دلالة عملية لصانعى القرار.

الصدق التلازمى وطرق تقديره:

سبق أن ميزنا بين نوعين من الصدق المرتبط بمحك أحدهما الصدق التنبؤى Predictive Validity الذى ناقشناه فى الجزء السابق، والصدق التلازمى Concurrent Validity، وذكرنا أن الصدق التنبؤى يتعلق بدرجة اقتران تباين درجات اختبار تنبؤى بتباين درجات اختبار محك يطبق بعد مرور مدة زمنية معينة، تفصل بين تطبيق كل منهما.

أما الصدق التلازمى فيتعلق بدرجة اقتران تباين درجات اختبار بتباين درجات اختبار آخر يطبق فى الوقت نفسه تقريبا، أى أن الصدق التنبؤى يهتم بالتنبؤ، أما الصدق التلازمى فيهتم بالوصف. فمثلاً يمكن مقارنة درجات الاختبارات أو المقاييس التى تطبق بهدف التشخيص الكلينيكى مع أحكام الأطباء النفسيين الذين يجرون مقابلات مع المرضى، أو يمكن مقارنة الكفاية فى مهنة معينة بتقديرات المشرفين على العمل، كذلك يمكن للمعلم أن يقارن درجات اختبار تحصيلى مقنن فى مجال دراسى معين بدرجات اختبار تحصيلى يعده بنفسه لتلاميذه فى المجال المعين، أو يمكن مقارنة



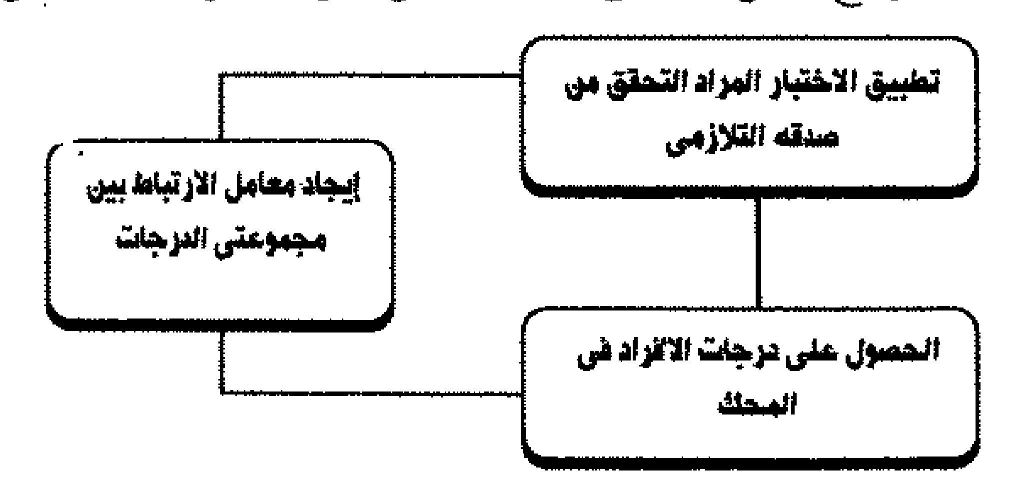
درجات اختبار جديد تم بناؤه لقياس سمة معينة بدرجات اختبار معترف به، أي تم التحقق من صدقه بأدلة متعددة. ففي هذه الحالات يتركز الاهتمام على الصدق التلازمي الذي يهدف لتحديد ما إذا كان من الممكن الاستعاضة عن درجات أحد الاختبارات بدرجات اختبار آخر.

وهذا يتطلب أن تكون هناك علاقة بين الاختبارين، وأن يكون الاختبار البديل متاحًا، ويتميز باتساق درجاته وسهولة تطبيقه بكلفة منخفضة. فالصدق التلازمي إذن يفيد في إحلال اختبار أو مقياس محل آخر، أما الصدق التنبؤي فيستخدم درجات اختبار أو مقياس معين في التنبؤ بدرجات اختبار محك؛ ولذلك فإنهما يمثلان جانبان متكاملان للصدق المرتبط بمحك.

ويستخدم الصدق التلازمي بدرجة أكبر في مجالات الصناعة وانتقاء الأفراد والتشخيص الكلينيكي، فمثلا يمكن تحليل العمل Job Analysis لتحديد المهارات أو السلوك الذي يتطلبه، وتقدير درجة تملك الفرد الذي يود الالتحاق بهذا العمل لعينات من السلوك ممثلة لهذه المهارات.

وقد نال هذا الأسلوب اهتمام المشتغلين بعلم النفس الصناعي في مجال انتقاء العاملين في المهن المختلفة، حيث بينت دراساتهم ارتباط مثل هذه العينات من السلوك بالأداء الفعلي في العمل ارتباطًا مرتفعًا. لذلك تستخدم هذه العينات السلوكية في فحص وانتقاء العاملين في مهن معينة استنادًا إلى دراسات الصدق التلازمي.

ويوضح الشكل التخطيطي (٥ ـ ٤) التالي خطوات تقدير الصدق التلازمي:



شكل (٥ ـ ٤) يوضح خطوات تقدير الصدق التلازمي للاختبار

إذ يمكن التحقق من الصدق التلازمي بتطبيق اختبار مناسب مثل اختبار كفاية، أو اختبار تحصيلي، أو اختبار معلومات متعلقة بعمل معين على مجموعة من الأفراد العاملين بالفعل، وفي الوقت نفسه تقريبًا نطبق اختبار محك يقيس مهاراتهم الفعلية في العمل، ونقدّر الصدق التلازمي بإيجاد قيمة معامل الارتباط بين درجات كل من الاختبارين.

كذلك يستخدم الصدق التلازمي عندما نطبق اختبار معين على أفراد تم بالفعل تشخيصهم تشخيصًا كلينيكيًّا، وذلك لمعرفة ما إذا كانت متوسطات درجات المجموعات المتباينة تختلف فيما بينها اختلافا دالا. فإذا بينت النتائج اختلاف قيم متوسطات الدرجات اختلافًا كبيرًا من مجموعة إلى أخرى، فإن هذا الاختبار يمكن أن يكون بديلاً مناسبًا لأغراض تصنيف الأفراد.

فمثلاً استبيان الشخصية متعدد الأوجه (MMPI) يُعد مقياسًا جيدًا لتصنيف الاضطرابات العقلية، حيث بينت الدراسات المتعلقة بهذا المقياس أن كثيرًا من حالات الاضطرابات العقلية التي يتم تشخيصها بواسطة الأخصائيين النفسيين تتميز درجاتهم التي يحصلون عليها في مجموعات معينة من مفردات المقياس (بعض المقاييس الفرعية) بنمط معين.

كذلك يمكن أن يقدِّم الصدق التلازمي معلومات مفيدة للمعلم تتعلق بالاختبارات الصفية التي يقوم بإعدادها، أو الاختبارات التحصيلية المقننة، ومدى مواءمتها للأهداف التعليمية التي يهتم بتحقيقها لدى تلاميذه. إذ يمكن أن يتحقق المعلم مما إذا كانت الاختبارات التحصيلية التي يُعدها في مجال دراسي معين تميز بالفعل بين الطلاب الأقوياء والطلاب الضعاف، وذلك بإيجاد الارتباط بين درجات كل اختبارين على حدة، أو بين الدرجات الكلية للاختبارات معًا وبين درجات اختبار تحصيلي مقنن يفي بخصائص الاختبارات الجيدة ويتعلق بالمجال الدراسي نفسه. كما يمكنه أن يقدر العلاقة بين مجموعة الدرجات باستخدام الشكل الانتشاري الذي أوضحناه عند مناقشتنا لمفهوم الارتباط في الفصل الثالث. فعندئذ يمكنه تمثيل درجات الاختبار التحصيلي المقنن على المحور الأفقي، ودرجات الاختبار الذي أعده بنفسه على المحور الرأسي، ووضع علامات تقابل أزواج درجات كل من الاختبارين لمجموعة الطلاب.

ويمكن توضيح ذلك لدرجات افتراضية بشكل (٥ ـ ٥) التالي:

درجات اختبار التحصيل المقنن	٤ـ٩	١٠ـ١٥	١٦ـ٢١	٢٢ـ٢٧	٢٨ـ٣٣
٣٠ـ٣٤				XXX X	XXX XX
٢٥ـ٢٩			XX	XXXX XX	XXX
٢٠ـ٢٤		• XX	XXXX XXXX		
١٥ـ١٩	X	X	XX		
١٠ـ١٤	XX	XXX			
٥ـ٩	X				

درجات الاختبار الذى أعده المعلم

شكل (٥ ـ ٥) يوضح انتشار درجات اختبارين تحصيليين

فمعظم العلامات فى الشكل ينبـغى أن تقع فى الربع الأدنى الأيسر الذى يحتوى درجات الأفراد الضعاف فى كل من الاختـبارين، وكذلك فى الربع الأعلى الأيمن الذى يحتـوى درجات الأفراد الأقوياء فى كل من الاخـتبارين، أمـا بقية الدرجات فـتقع فى الربعين الباقيين.

ولكن النزعة العامـة للعلاقات ينبغى أن تتجـه من الربع الأدنى الأيسر إلى الربع الأعلى الأيمن لكى يدل على نزعة موجبة مما يقدم مـؤشراً جيداً للصدق التلازمى. وإذا لم يحدث ذلك فإن الصدق التلازمى يكون منخفضاً، وعندئذ ينبغى على المعلم أن يعيد النظر فى الاختبار الذى أعده، أو يتحقق من أن الاختبار التحصيلى المقنن مناسباً ومرتبطاً بالأهداف التعليمية التى يود قياسها.

كما أن الصدق التلازمى يمكن أن يفيـد المعلم نظراً لصعوبة تقدير الصدق التنبؤى للاختبارات الصفية . فالمعلم يهتم عادة بالسلوك المستقبلى لتلاميذه ، وبالطبع لا يستطيع إجراء دراسات تتبعيـة بعيدة الأمد لتقدير مدى صدق الأساليب التعليمـية التى يتبعها ؛ لذلك يمكنه تقـدير صدق هذه الأساليـب من خلال سلوك تلاميـذه فى الوقت الراهن. فـمن المعلـوم أن درجـة العـلاقـة بين متـغيـرين تقـل بمرور الزمـن ؛ لـذلـك ترتبط

اختبارات الاستعدادات بالتحصيل الدراسى فى الوقت الراهن بدرجة أكبر من ارتباطها به بعد مرور عدة سنوات.

لذلك نتوقع انخفاض قيمة الصدق التنبؤى عن قيمة الصدق التلازمى، فقيم معامل الصدق التلازمى تُعد قيمًا قصوى لمعامل الصدق التنبؤى، وهذا يعنى أنه إذا كانت قيمة معامل الصدق التلازمى ٦٥, ٠ مثلاً ، فإننا نتوقع أن يكون الصدق التنبؤى أقل من ذلك إذا استخدم اختبار المحك نفسه فى الحالتين.

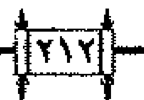
بعض العوامل التى تؤثر فى الصدق المرتبط بمحك :

تتأثر قيم معامل الصدق المرتبط بمحك بعوامل متعددة أشرنا إلى بعض منها فى سياق مناقشاتنا السابقة. وسوف نوضح فيما يلى مزيدًا من هذه العوامل التى ينبغى على مستخدمى الاختبارات والمقاييس والمسئولين عن اتخاذ قرارات الانتقاء والتسكين مراعاتها عند تفسير البيانات المتعلقة بهذا النوع من الصدق.

مدى تجانس عينة الأفراد: نظرًا لأن الصدق المرتبط بمحك يعتمد على معامل الارتباط بين درجات الاختبار التنبؤى واختبار المحك، فإن جميع العوامل التى تؤثر فى قيم معامل الارتباط ومعامل الثبات التى أوضحناها فى الفصلين الثالث والرابع تؤثر أيضًا فى قيم معامل الصدق. فكلما كانت المجموعة متجانسة انخفضت قيمة معامل الصدق المرتبط بمحك، وكلما كانت غير متجانسة ازداد تباين درجات الاختبار وبالتالى تزداد قيمة معامل الصدق. وتجانس المجموعة يعد من المشكلات المتعلقة بالصدق عند اتخاذ قرارات انتقاء الأفراد. فقيمة معامل الصدق تستمد عادة من تطبيق الاختبار التنبؤى على مجموعة من الأفراد الذين تم انتقاؤهم بالفعل، وهذه المجموعة تكون أكثر تجانسًا من المجموعة التى سيتخذ قرار الانتقاء بشأنها فى المستقبل مما يؤدى إلى انخفاض هذه القيمة مما قد يثير الشك فى فاعلية الاختبار التنبؤى.

لذلك ينبغى عند تقدير قيمة معامل الصدق التنبؤى لاختبار معين أن تُستخدم الدرجات المستمدة من مجموعة الأفراد التى تتقدم للالتحاق بدراسة أو بعمل معين قبل أن تُجرى عمليات الفحص والانتقاء.

فمثلاً نوجد العلاقة بين درجات اختبار الاستعداد الدراسى لجميع المتقدمين للالتحاق بالكليات ودرجاتهم أو تقديراتهم فى نهاية السنة الأولى الجامعية، ولكن



الكليات لا تسمح عادة بإلحاق جميع الطلاب، وكذلك لا تسمح مؤسسات العمل بإلحاق جميع المتقدمين للمهن أو الوظائف.

لذلك تعتمد دراسات الصدق التنبؤى على الأفراد المقبولين فقط، مما يؤدى إلى ضيق مدى درجات الاختبارات التنبؤية؛ نظرًا لأن هؤلاء الأفراد يمثلون مجموعة أكثر تجانسًا من المجموعة التى ستطبق عليها هذه الاختبارات فى المستقبل بغرض انتقاء أفضلهم. وهذا التناقض فيما يتعلق بالصدق يتلخص فى أن الاختبارات التى تستخدم فى اتخاذ قرارات الانتقاء ينبغى أن تكون قيمة معامل صدقها التنبؤى مرتفعة بالنسبة للمجموعة التى ستطبق عليها هذه الاختبارات مستقبلاً قبل بدء عملية الانتقاء. ولكن نظرًا لأن هذه القيم يتم تقديرها فى واقع الأمر استنادًا إلى المجموعة المنتقاة، فإننا نتوقع أن تكون منخفضة.

لذلك ينبغى عند تفسير قيم معامل الصدق المرتبط بمحك عدم التسرع فى الحكم بعدم صلاحية الاختبار التنبؤى، أو عدم الثقة فى فاعليته فى التنبؤ بالنجاح فى الدراسة أو الأداء فى العمل، إذ ربما يرجع انخفاض قيم معامل الصدق إلى قلة تباين الدرجات فى الاختبار نتيجة الاعتماد على المجموعة المتجانسة التى تم انتقاؤها، أو ربما يرجع ذلك إلى الثبات المنخفض لدرجات الاختبار التنبؤى أو اختبار المحك، أو إلى عدم استقامة العلاقة بين درجات الاختبارين، أو غير ذلك من العوامل التى تؤثر فى قيم هذا المعامل.

ثبات درجات المحك: سبق أن أوضحنا أن مشكلة ثبات درجات الاختبار التنبؤى ودرجات الاختبار المحك تعد من المشكلات الأساسية، وذكرنا كيفية تصحيح الأخطاء التى ربما تسهم فى انخفاض قيم معامل ثبات كل منهما.

فالمحكات تكون عادة تقديرات المشرفين على العمل، أو نتائج ملاحظات المعلم أو المعدل التراكمى للطالب، وجميعها يشوبها الخطأ مما يؤثر فى قيم معامل ثبات درجات المحك. كما سبق أن ذكرنا عند مناقشتنا لمفهوم الثبات أن قيمة معامل صدق الاختبار تكون دائمًا أقل من قيمة مؤشر الثبات، أى أقل من $\sqrt{ر_{٢٢}}$ حيث $\sqrt{ر_{٢٢}}$ ترمز إلى ثبات درجات المحك.

فمثلاً إذا كانت قيمة معامل الثبات ٠,٤٠ فإن معامل صدق المحك لا يمكن أن يزيد عن $\sqrt{٠,٤٠}$ أى ٠,٦٣ حيث إن درجات المحك لا ترتبط عادة بدرجات اختبار آخر أكثر من ارتباطها بنفسها.

تأثر المحك بمتغيرات أخرى: يتأثر معامل الصدق المرتبط بمحك بالمتغيرات الأخرى التى تتداخل مع المحك، فعندما نوجد معامل الارتباط بين درجات اختبار

الاستعداد الميكانيكى مثلا، وتقديرات المشرفين على العمل لمجموعة من الأفراد، فإن معرفة المشرف على العمل بدرجات الأفراد فى الاختبار ربما تؤثر فى تقديراته لهم، مما يؤثر فى قيمة هذا المعامل؛ لذلك لا ينبغى على المشرف الاطلاع على درجات اختبار الاستعداد عند تقديره للعاملين.

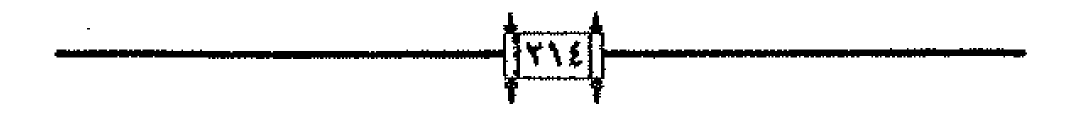
عدد مفردات الاختبار التنبؤى: نظراً لأن القيمة القصوى للصدق المرتبط بمحك تعتمد على مؤشر الثبات $\sqrt{r_{tt}}$ ، فإن هذه القيمة تزداد بزيادة قيمة معامل الثبات. وزيادة قيمة معامل الثبات تعتمد على طول الاختبار، أى عدد مفرداته، فكلما زاد هذا العدد زادت قيمة معامل الثبات إذا طبق الاختبار على مجموعة الأفراد نفسها. وقد سبق أن أوضحنا كيفية تقدير قيمة معامل الثبات إذا أطلنا الاختبار، أى أضفنا إليه بعض المفردات التى تقيس السمة نفسها، وذلك باستخدام الصيغة العامة لسبيرمان وبراون Spearman - Brown . ويمكن الرجوع إلى جدول (٤ ـ ١) لمعرفة تأثير هذه الإطالة على القيم المتوقعة لمعامل الثبات وبخاصة القيم المتوسطة، غير أن هذا التأثير يكون أقل فيما يتعلق بقيم الصدق.

المدة الزمنية الفاصلة بين تطبيق الاختبار التنبؤى واختبار المحك: فمما لا شك فيه أننا نتوقع انخفاض قيمة معامل الصدق المرتبط بمحك بزيادة المدة الزمنية التى تفصل بين تطبيق كل من الاختبارين، نظراً لتأثير الأخطاء العشوائية فى درجات كل من الاختبارين، ولذلك يفضل أن تكون هذه المدة الزمنية قصيرة لكى يتيسر إيجاد معامل الارتباط بين درجاتهما، ويتحقق هذا بوضوح فى الصدق التلازمى.

عدد أفراد مجموعة المختبرين : فكلما قل عدد الأفراد الذين يطبق عليهم الاختبار التنبؤى نتوقع عند تطبيق المحك بعد مرور مدة زمنية معينة أن يقل العدد أكثر مما يؤثر كما ذكرنا فى قيمة معامل الصدق التنبؤى. لذلك ينبغى البدء بعينة كبيرة من الأفراد حيث يترتب على طول المدة الزمنية فقدان بعضهم نتيجة الانقطاع أو الرسوب أو أى ظروف طارئة أخرى.

صدق التكوين الفرضى (صدق المفهوم) :

يعد مصطلح الصدق Validity من أكثر مصطلحات القياس التربوى والنفسى إثارة للجدل وتباين وجهات النظر، والدليل على ذلك تعدد المقالات المتعلقة به والمنشورة فى الدوريات المتخصصة التى تصدر عن الجمعية الأمريكية لعلم النفس (APA)، والجمعية الأمريكية للبحث التربوى (AERA) ، والجمعيات والمجالس القومية المتخصصة فى القياس التربوى والنفسى.



وتهتم هذه المقالات بمراجعة مصطلح الصدق ومحاولة توسيع مضمونه لكى يأخذ فى الاعتبار ما يترتب على استخدامات أساليب القياس فى المجالات النفسية والتربوية، وبخاصة فى مجال قياس السمات الإنسانية. فصدق التكوين الفرضى أو صدق المفهوم Construct Validity نشأ وتطور نتيجة النقد المستمر للمفهوم التقليدى للصدق. فالذى يطلع على البحوث التربوية والنفسية يجد كثيرًا من التسميات المتعلقة بصدق أدوات القياس، مثل الصدق الظاهرى، والصدق المنطقى، والصدق التجريبى، وصدق المضمون، والصدق المنهجى، والصدق الذاتى، وصدق التعميم ، والصدق العاملى، والصدق الإحصائى، والصدق التمايزى، والصدق التطابقى، والصدق التقاربى، وغيرها مما يثير العديد من التساؤلات حول مصطلح الصدق وفيم يستخدم.

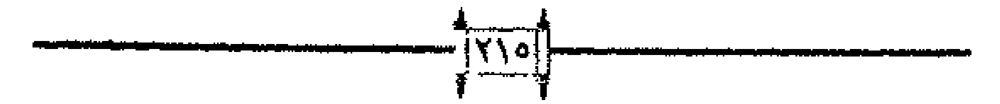
ولعل هذا ما دعا هذه الجمعيات إلى مراجعة هذا المصطلح، وأوصت بتقسيم دراسات الصدق إلى الصدق المرتبط بمحك بجانبيه التنبؤى والتلازمى، وصدق المحتوى، وصدق التكوين الفرضى، أو صدق المفهوم.

وتتعلق هذه الدراسات بأنواع القرارات الثلاثة التى أشرنا إليها فى مستهل الجزء الخاص بالصدق. وعلى الرغم من أن مصطلح صدق التكوين الفرضى قد اقترحه كل من كرونباك Cronbach وميهل Meehl عام ١٩٥٠، إلا أنه قد نال اهتمامًا ملحوظًا وقبولاً متزايدًا من جانب علماء القياس وبخاصة فى الآونة الأخيرة نتيجة غموض كثير من التكوينات الفرضية أو المفاهيم السيكولوجية، الذى ترتب عليه إعاقة تطوير اختبارات ومقاييس أكثر صدقًا.

ويتناول صدق التكوين الفرضى العلاقة بين نتائج الاختبارات والمقاييس وبين المفهوم النظرى الذى يهدف الاختبار لقياسه، مثل مفاهيم الذكاء، والقلق، وقوة الأنا، والتصلب، والانطواء، ودافعية الإنجاز، والابتكارية، والاستعداد المكانى، والقدرة التحليلية، ومستوى الحافز ، والمناخ المدرسى، والإدارة الذاتية، والقيادة المدرسية، وهكذا.

وبعبارة أخرى فإن صدق التكوين الفرضى يهدف لتحديد التكوينات الفرضية التى يُعزَى إليها تباين الأداء فى الاختبارات، أى أن هذه التكوينات الفرضية هى التى يتركز عليها الاهتمام وليس درجات اختبار المحك أو سلوك الفرد.

ولكن ما المقصود بالتكوينات الفرضية، وهل تختلف عن السمات الإنسانية؟ فعلى الرغم من أن كرونباك Cronbach استخدم المعنى العام أو القاموسى لمصطلح السمات Traits من حيث إنها خصائص موجودة داخل الفرد، وكذلك المعنى العام لمصطلح التكوينات الفرضية Hypothetical Constructs من حيث إنها تتواجد فى



فكر علماء النفس، إلا أن لوفنجر (Loevinger, 1967) قارنت بين المصطلحين بما يشبه المقارنة التى تشير إليها المراجع الإحصائية بين بارامتر أو معْلم المجتمع Parameter وإحصاء العينة Statistic.فبارامتر المجتمع هو الذى يهدف علم الإحصاء الاستدلالى إلى تقدير قيمته، أما إحصاء العينة الذى يناظر البارامتر فإنه يمثل أفضل تقدير متاح لهذا البارامتر.

ولذلك فإن السمات هى التى نهدف لفهم طبيعتها كما أوضحنا فى الفصل الأول، أما التكوينات الفرضية المناظرة لهذه السمات، فإنها تمثل أفضل ما يتوصل إليه العلماء فى وقت ما لتوضيح طبيعة هذه السمات. ولذلك يركز صدق التكوين الفرضى على ثلاثة عناصر هى : الاختبار، والسمات المراد قياسها، وماذا يقيس الاختبار من وجهة نظر القائم بإعداده، أى أن هذه العناصر تتعلق بالتكوين الفرضى، والتفسير، والنظرية التى يستند إليها التفسير. ويرى كرونباك أن هناك تساؤلين مستقلين يتضمنهما صدق التكوين الفرضى هما: إلى أى حد يقيس الاختبار ما يتضمنه تكوين فرضى أو مفهوم معين؟ وإلى أى حد يتضمن المفهوم فرضًا صادقًا؟

غير أن لوفنجر Loevinger تصوغ التساؤلين صياغة بديلة كالتالى:

إلى أى حد يقيس الاختبار سمة لها وجود فعلى؟ وإلى أى حد يكون هناك تناظر بين التفسير المقترح وما يقيسه الاختبار؟

فالتساؤلان اللذان اقترحهما كرونباك يقسّمان صدق التكوين الفرضى إلى: صدق الاختبار فيما يتعلق بتكوين فرضى معين، وصدق التكوين الفرضى ذاته، ويرى أن هذين النوعين من الصدق لا ينفصلان عادة إلا إذا كان الاختبار أو التكوين الفرضى قد نال قبولاً واسعًا على مدى زمنى طويل.

أما التساؤلان الآخران فيقسّمان صدق التكوين الفرضى إلى الصدق الذاتى للاختبار، وصدق التفسير. وترى لوفينجر أن هذه الصياغة البديلة تسمح بتقييم التساؤلين كل على حدة، فقيم الارتباط الداخلى بين مفردات الاختبار، والقيم المرتفعة للارتباط بين درجات الاختبار ومتغيرات خارجية تقدم أدلة على أن الاختبار يقيس شيئًا ما بطريقة منظمة، أما محتوى المفردات وطبيعة العلاقات القائمة بينها، وطبيعة المتغيرات الخارجية التى تتباين فى درجة ارتباطها بالاختبار فتقدم أدلة على التفسير المقترح. أى أن الأدلة الأولى تفيد فى إضفاء معنى سيكومترى على الاختبار، أما

الأدلة الثانية فتُضفي على الاختبار معنى سيكولوجي أو تربوى، وكلاهما أمرًا ضروريا. فالعلم يجب أن يُقدم إطارًا نظريًا يسمح بتكامل وتفسير البيانات المتعلقة بمفهوم معين، وتوجيه البحوث المستقبلية المتعلقة به.

طبيعة التكوينات الفرضية :

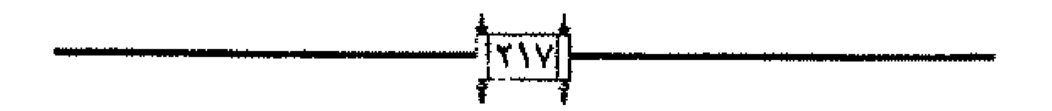
إن التكوينات الفرضية Constructs تعد بمثابة خصائص يُفترض أنها تُميز الأفراد، وينعكس أثرها فى سلوكهم. ويمكن أن تكون هذه التكوينات كيفية مثل فقدان الذاكرة أو كمية مثل الذكاء، ويفترض أن الفرد الذى يمتلك خاصة معينة يحتمل أن يسلك فى موقف معين سلوكًا معينًا.

وعادة يجرى تعريف هذه التكوينات الفرضية تعريفًا إجرائيًا بحيث يمكن قياس أثرها بواسطة أدوات القياس. ولكن يمكن أيضًا أن يتخطى التعريف السلوك الملاحظ بحيث يكون أكثر اتساعًا ودلالة من مجرد إعادة صياغة العلاقات الإمبيريقية فى التعريف الإجرائى. ويمكن كذلك أن تكون التكوينات الفرضية عيانية Concrete مثل زمن الرجع، أو تكون مجردة Abstract مثل الشخصية أو الدافعية، أو الطلاقة اللفظية، أو الذكاء، أو التوافق ، أو تقدير الذات، أو القيادة التربوية.

فكلما كان التكوين الفرضى أو المفهوم عيانيا أمكن بسهولة التحقق من صدق المقياس المتعلق به، والعكس إذا كان المفهوم مجردًا. فدرجة صعوبة التحقق من صدق التكوين الفرضى تتناسب ودرجة عيانية أو تجريد المفهوم المراد قياسه. فمثلا يمكن بسهولة قياس مفهوم زمن الرجع Reaction Time وذلك عن طريق حساب الزمن الذى يستغرقه الفرد فى استجابته بالضغط على زر جرس عند سماعه أو رؤيته مثيراً معينًا.

فالمتغير موضع الاهتمام فى هذه الحالة هو سرعة الاستجابة التى يمكن قياسها بسهولة. وتعتمد دقة القياس على التقنية المستخدمة فى تقديم المثير وتسجيل الزمن. أى أننا هنا نهتم اهتمامًا مباشرًا بإجراءات القياس، ولا نكون فى حاجة إلى تقدير صدق المقياس، ومن ثم يمكن إجراء دراسات لتحديد علاقات معينة بين المقياس ومتغيرات أخرى.

ولكن كثيرًا من المفاهيم السيكولوجية ليست بهذا القدر من البساطة، حيث إنها مفاهيم مجردة تمتد أبعد من حدود إجراءات قياس معينة. فمفهوما الدافعية Motivation وقوة الأنا Ego Strength يمكن أن نستدل عليهما من أنماط متعددة من السلوك مما



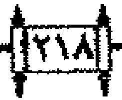
يتطلب التحقق من صدق المقاييس التى نفترض أنها تقيس المفهوم المجرد، وهذا بالطبع يتطلب تصوراً فكريا معيناً فيما يتعلق بالمفهوم وأبعاده المتكاملة. أى أن المفهوم أو التكوين الفرضى يُمثل فرضاً يشير إلى ارتباط أنماط متعددة من السلوك فيما بينها فى دراسة الفروق الفردية.

لذلك يرى نانلى (Nunnally , 1978) ، وزيلر (Zeller, 1978) أن جميع النظريات العلمية تهتم اهتماماً أساسياً بدراسة التكوينات الفرضية ولا تهتم كثيراً بالحالات الخاصة، أى تهتم بالعموميات دون الخصوصيات. فالعلماء لا يستطيعون إجراء بحوثهم بدون هذه التكوينات الفرضية والتعبير عنها والتواصل بهما، ويسعون دوماً إلى تطوير مقاييس لهذه التكوينات والتوصل إلى علاقات وظيفية بين مقاييس التكوينات الفرضية المختلفة. فصدق التكوين الفرضى يعد من المشكلات الأساسية فى التعميمات العلمية، فدراسة تكوين فرضى معين مثل الذكاء أو معدل التعلم يتطلب استخدام مقياس أو أكثر بحيث تتميز نتائجها بقابلية التعميم على فئة من المقاييس التى تتعلق بالتكوين الفرضى نفسه.

ولذلك فإن الأسس المنطقية التى تربط صدق التكوين الفرضى بالتعميمات العلمية تعد أكثر تعقيداً من الأسس التى يعتمد عليها الصدق المرتبط بمحك.

وتتباين التكوينات الفرضية فى درجة اتساع وتحديد نطاق السلوك الذى يستدل منه عليها. فأحياناً يكون هذا النطاق محدوداً حيث يشتمل على عدد قليل من المتغيرات التى يمكن ملاحظتها وقياسها مثل زمن الرجع الذى أشرنا إليه، وأحياناً كثيرة يكون النطاق متسعاً بل مفتوحاً بحيث يشمل العديد من المتغيرات المتداخلة مما يلقى بظلال من الغموض على التكوين الفرضى. ومع هذا يحاول علماء النفس تحديد بعض المتغيرات الأساسية المتعلقة بالتكوين الفرضى تحديداً واضحاً، ويصوغون افتراضات تتعلق بالمتغيرات الأخرى التى ربما يرون أن التكوين الفرضى يمكن أن يشملها. ولتوضيح ذلك نجد أن علماء النفس قد اتفقوا على أن قياس مفهوم الذكاء ينعكس فى سلوك حل المشكلات التى تتطلب قدرات استدلالية، ولكنهم اختلفوا فيما إذا كانت القدرات الحسية الإدراكية والذاكرة تعد جزءاً من هذا المفهوم.

وينطبق ذلك على كثير من المفاهيم السيكولوجية والتربوية، حيث تكون حدود نطاق السلوك الملاحظ الذى يستدل منه على المفهوم غير واضحة وضوحاً تاما، لذلك يفضل قياس المفاهيم أو التكوينات الفرضية استناداً إلى ضم نتائج مقاييس متعددة لأنماط السلوك المتعلقة بها بدلا من الاعتماد على كل مقياس على حدة من أجل زيادة صدق التعميمات العلمية.



ولكن نظراً لما يتطلبه ذلك من جهد كبير، فإن الاعتماد على مقياس واحد لمفهوم معين يجعل هناك درجات من صدق التكوين الفرضي بحسب اقتراب نتائج هذا المقياس من نتائج مقاييس أخرى للمفهوم ذاته. وهذا يدل على أن تقييم صدق التكوين الفرضي يتطلب تجميع أدلة متعددة من مصادر متباينة وفي مواقف مختلفة حيث إنه لا يوجد مؤشر كمي لهذا النوع من الصدق.

منهجية التحقق من صدق التكوين الفرضي:

إن عملية التوصل إلى أدلة عن صدق التكوين الفرضي يمكن اعتبارها محاولة للإجابة عن التساؤلين السابقين اللذين اقترحهما كرونباك Cronbach أو لوفنجر Loevinger، وهذا يتطلب كما ذكرنا جمع أدلة من مصادر متعددة وبطرق مختلفة وتقييم هذه الأدلة تقييمًا متكاملاً. إذ يمكن فحص مفردات الاختبار لمعرفة المثيرات التي تشتمل عليها ومحتواها ونوعها، ومدى استقرار درجات الاختبار في ظروف متباينة، ومدى تجانس مفرداته، وارتباط درجاته بدرجات اختبارات أخرى وبمتغيرات ربما تختلف عما يقيسه الاختبار، ومدى تمييزه بين مجموعات مختلفة من الأفراد فيما يقيسه الاختبار، وتأثير الظروف التجريبية على أداء المختبرين، وغير ذلك من البيانات التي تلقي الضوء على ما تدل عليه درجات الاختبار.

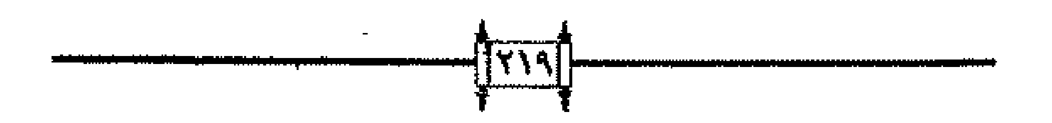
وقد اقترح كرونباك Cronbach وميهل Meehl في مقالتهما الرائدة في هذا المجال عام ١٩٥٥ توجهًا مختلفًا في هذا الشأن، حيث أشارا إلى منظور آخر لعملية جمع الأدلة عن صدق التكوين الفرضي يتركز حول بناء نظرية مصغرة Miniature Theory تتعلق بالاختبار في ضوء الخطوات الثلاث التالية:

(١) تبنى نظرية معينة يستند إليها الاختبار، وتُشتق منها فروضًا معينة، تتعلق بالسلوك المتوقع من الأفراد الذين يحصلون على درجات مختلفة في الاختبار.

(٢) جمع بيانات يستند إليها في التحقق من صحة الفروض.

(٣) تقييم ما إذا كانت النظرية تفسر بالفعل البيانات المتجمعة تفسيراً كافيًا، فإذا لم يكن الأمر كذلك ينبغي مراجعة النظرية، وتكرر الخطوات السابقة للتوصل إلى تفسير أكثر كفاية.

وعلى الرغم من أن إجراء هذه الخطوات يبدو أمراً يسيراً إلا أنها في واقع الأمر تتطلب جهداً كبيراً ودراسات متعددة وأحكامًا تقييمية تستند إلى نتائج بحوث تجريبية.



لذلك ربما يكون من المناسب أن نقدم مزيداً من التوضيح لهذه الخطوات فى إطار مقترحات كرونباك Cronbach وميهل Meehl وبراون Brown. فالنظرية التى يستند إليها الاختبار تشتمل على مجموعة من المفاهيم والقضايا والمبادئ المترابطة، لذلك يطلق عليها « شبكة القوانين المترابطة Monological Network ».

وتفيد هذه الشبكة فى الربط بين الخاصة أو السلوك الملاحظ وبين التكوين الفرضى النظرى، أو بغيره من التكوينات الفرضية، أو بخاصة ملاحظة أخرى، أى أن هذه الشبكة بما تتضمنه من قوانين تعبِّر عن علاقات محددة أو احتمالية.

وهنا تأتى عملية الاستدلال على معنى التكوين الفرضى من البيانات التى تمثل السلوك الملاحظ. وهذه العملية يجب أن تكون محددة تحديداً دقيقاً بحيث يمكن لباحثين مختلفين التحقق منها والتواصل بها. وهذا القوانين ينبغى أن تكون ذات صلة بسلوك ملاحظ فى وقت ما. أما إذا كانت القوانين مشتقة من قوانين أخرى، فإن هذه القوانين الأخرى هى التى ينبغى أن تفى بهذا الشرط. ويمكن التعبير عن هذا السلوك الملاحظ كيفيًّا مثل متفائل أو مرح، أو كميًّا مثل درجات الاختبارات، فكلاهما يفيد فى قياس التكوين الفرضى.

وهذا لا يتأتى إلا إذا استخدم هؤلاء الباحثين شبكة القوانين المترابطة المقترحة والمتعلقة بتكوين فرضى معين، وتتضمن عملية الاستدلال التحقق من فروض معينة فى ضوء هذه الشبكة واستناداً إلى البيانات التى يتم جمعها.

وينبغى توضيح تعريف التكوين الفرضى المراد التحقق من صدقه، والكيفية التى اشتقت بها الفروض من النظرية المتبناة، والإجراءات التجريبية، والأسلوب الذى اتبع فى الاستدلال على معنى التكوين الفرضى المعين من هذه الإجراءات، وتقييم الأدلة التى أكدت صحة الفروض، والأدلة الواهنة.

ولعل المشكلة التى تواجه القائمين بذلك تتعلق بهذه الأدلة الواهنة، حيث إنه ربما يصعب تحديدها بدقة وتبريرها؛ إذ ربما يكون الاختبار مقياساً غير مناسب للتكوين الفرضى المعين، أو أن النظرية التى استند إليها يشوبها بعض أوجه القصور، أو أن التصميم التجريبى أو إجراءاته غير مناسبة للتحقق من صحة الفروض.

لذلك يرى كرونباك Cronbach وثورنديك Thorndike وأنستازى Anastasi أن هذه الخطوات مثالية وربما لا تتحقق بدرجة كافية فى كثير من التكوينات الفرضية أو المفاهيم السيكولوجية والتربوية، ويقترحون محاولة الاقتراب من هذه المثالية بأن

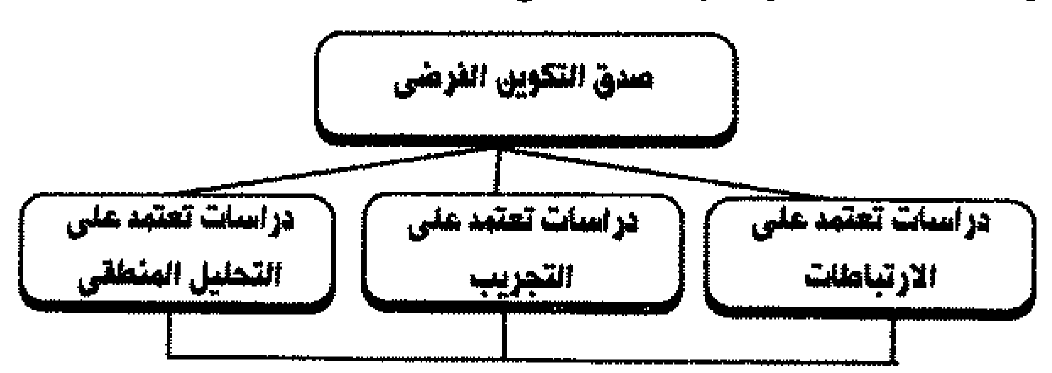
يستخدم الاختبار كمؤشر Signبحيث يقدم معلومات توضح طبيعة النطاق السلوكي المتعلق بمفهوم أو تكوين فرضي معين، وبالتالي يعد مؤشراً لطبيعة هذا المفهوم أو هذا التكوين الفرضي.

أساليب جمع أدلة تتعلق بصدق التكوين الفرضي :

اتضح مما سبق أن صدق التكوين الفرضي يتطلب جمع معلومات وأدلة من مصادر متعددة، وفحص نتائج الدراسات المتعلقة بمفهوم أو تكوين فرضي، معين والربط فيما بينهما.

فالنظرة المتكاملة لهذه الأدلة تلقي الضوء على طبيعة التكوين الفرضي،والعوامل التي تسهم في الكشف عنه وتطويره؛ لذلك ربما لا نعجب إذا لاحظنا تعدد أساليب جمع هذه الأدلة ، نظراً لأن البحث في صدق التكوين الفرضي يعد عملية مستمرة وفقاً لما يطرأ على تكوين فرضي معين من متغيرات، أو ما يتوصل إليه الباحثون من نتائج ربما تخالف النتائج التي اعتقد البعض أنها أكدت صدق هذا التكوين الفرضي.

غير أننا سوف نركز مناقشتنا على بعض الأساليب الأساسية المستخدمة في دراسات الصدق، وسوف نوضحها بالشكل التخطيطي (٥ ـ ٦) التالي:



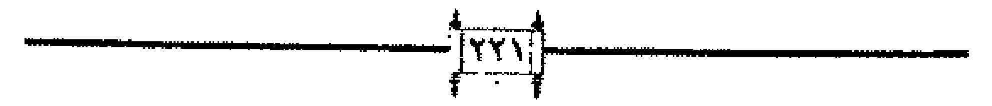
شكل (٥ ـ ٦) يوضح بعض الأساليب المستخدمة في دراسات صدق التكوين الفرضي

ويتضح من شكل (٥ ـ ٦) أنه يمكن تصنيف هذه الأساليب في ثلاث مجموعات متكاملة هي:

(١) دراسات تعتمد على الارتباطات.

(٢) دراسات تعتمد على التجريب .

(٣) دراسات تعتمد على التحليل المنطقي.



ومما لا شك فيه أن هذه الأساليب يغلب عليها الطابع التكاملي، فكل منها يثرى الآخر ويسترشد بنتائجه. فالدراسات الارتباطية تهتم بتحديد طبيعة الفروق بين الأفراد الذين يحصلون على درجات مرتفعة فى الاختبار والأفراد الذين يحصلون على درجات منخفضة، وكذلك تحليل العلاقات القائمة بين درجات مجموعة من الاختبارات للكشف عن السمات التى تشترك هذه الاختبارات فى قياسها واختلافها عن غيرها من السمات.

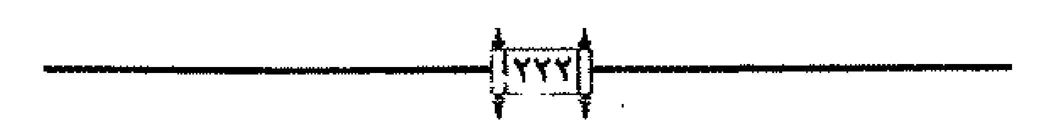
أما الدراسات التجريبية فتهدف لمعرفة التغيرات التى تحدث فى أداء الفرد فى اختبار معين نتيجة التأثير التجريبى فى أحد المتغيرات المتعلقة بالفرد للتحقق من صحة فرض معين يتعلق بما يقيسه الاختبار، بينما تركز الدراسات المنطقية على محتوى الاختبار وأساليب تصحيحه أو تقدير درجاته والعوامل التى تؤثر فى هذه الدرجات. وسوف نوضح فيما يلى كل من هذه الأساليب بمزيد من التفصيل.

أساليب تعتمد على الارتباطات :

تتعدد الأساليب التى تستخدم فى دراسات صدق التكوين الفرضى استنادًا إلى الارتباطات، فأحد هذه الأساليب يتعلق بدراسة الفروق بين مجموعات مختلفة من الأفراد Group Differences Method . وتهدف كثير من هذه الدراسات للمقارنة بين مجموعتين مختلفتين فى ضوء الدرجات التى يحصل عليها الأفراد فى اختبار أو مقياس معين، حيث يتوقع أن المجموعتين سوف تختلف استجاباتهما أو الدرجات التى يحصل عليها أفراد كل منهما فى الاختبار أو المقياس، ويمكن التحقق من ذلك بطريقة مباشرة.

فإذا أردنا التحقق مثلا من صدق التكوين الفرضى لاختبار يقيس العصابية Neuroticism ، فإنه يمكن اختيار مجموعتين من الأفراد استنادًا إلى تشخيص الأطباء النفسيين بحيث تكون هذه السمة مرتفعة لدى إحدى المجموعتين ومنخفضة لدى المجموعة الأخرى، ويطبق الاختبار بعد ذلك على كل من المجموعتين. فإذا تبين أن هناك اختلافًا واضحًا بين درجات كل من المجموعتين، فإنه يمكن اعتبار ذلك أحد أدلة صدق التكوين الفرضى للاختبار، ويستبقى عندئذ على المفردات التى ميزت بدرجة أكبر بين المجموعتين .

ويشير براون (Brown, 1972) فى هذا الشأن إلى أن النتائج التى نحصل عليها من مقياس للشخصية مثلا ربما تعزى إلى الخاصية أو السمة المقاسة، أو إلى طريقة القياس (استبيان، قائمة ملاحظة، بطاقة تقرير ذاتى، مقابلة أو غير ذلك)، أو إلى كليهما معًا.



فإذا درسنا السمة نفسها أو سمات متباينة بطرق مختلفة، فإنه ربما نحصل على نتائج مختلفة، لذلك إذا أردنا عزل الإسهام النسبي لكل من السمة أو الطريقة يجب دراسة أكثر من سمة وأكثر من طريقة معًا، أى ندرس الصدق التقاربي والصدق التمايزى. ففى الصدق التقاربى ننظر إلى الارتباطات بين السمات نفسها إذا أجرى قياسها بطرق مختلفة، وفى الصدق التمايزى ننظر إلى الارتباطات بين سمات متباينة إذا أجرى قياسها بطريقة واحدة.

والأسلوب الثانى الذى يستند إلى الارتباطات يعتمد على إيجاد معامل الارتباط بين درجات اختبار ويفترض أنه يقيس تكوينًا فرضيًا معينًا، ودرجات اختبار آخر بينت الأدلة المتعددة أنه يقيس التكوين الفرضى ذاته.

فمثلاً كثير من اختبارات الذكاء الجديدة التى يقوم الباحثون أو المؤسسات المتخصصة ببنائها تعتمد على إيجاد ارتباط درجات الاختبار الجديد بدرجات اختبار ستانفورد بينيه للذكاء Stanford - Binet، أو اختبار ويكسلر للذكاء Wechsler وهما اختباران شائعا الاستخدام ونالا كثيراً من دراسات الصدق على مدى زمنى طويل.

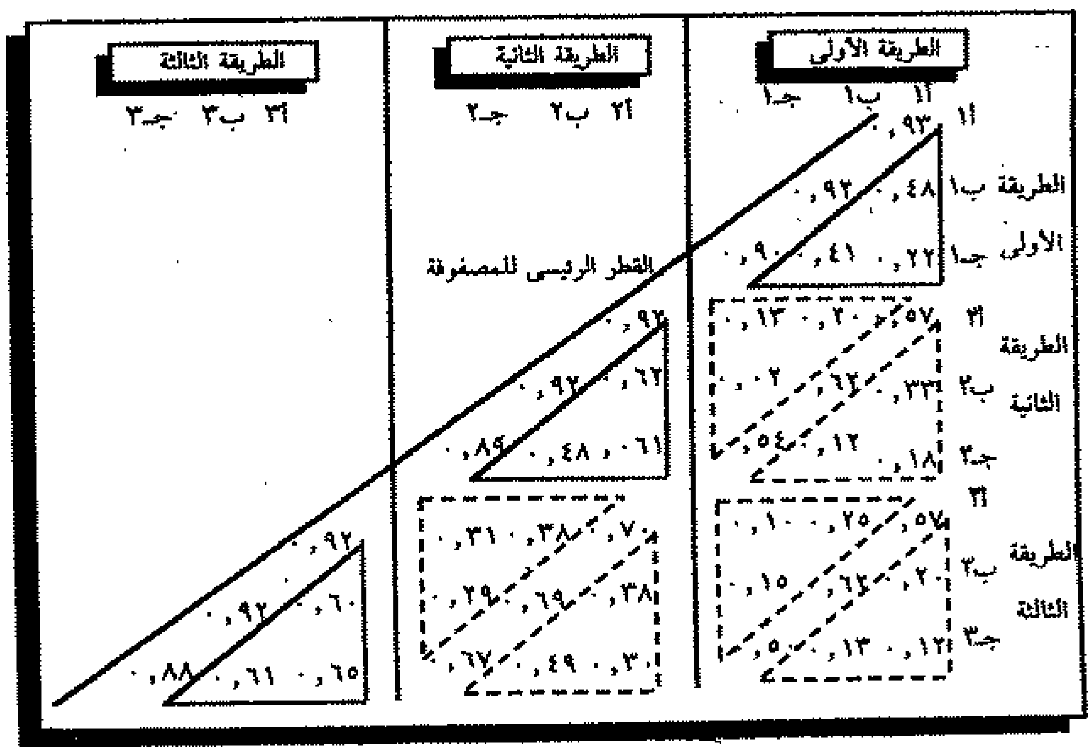
فإذا وجد أن معامل الارتباط قيمته مرتفعة فإن هذا يعد أحد أدلة صدق التكوين الفرضى. ويمكن كذلك أن نستدل على أن درجات الاختبار الجديد يمكن أن ترتبط بمتغيرات أخرى بنفس قدر ارتباط كل من الاختبارين المذكورين بهذه المتغيرات.

وعادة يمتد هذا الأسلوب بحيث يتناول العلاقات القائمة بين الاختبار ومجموعة أخرى من الاختبارات بدلا من اختبار واحد. ويمكن إجراء ذلك باستخدام أحد الأساليب الإحصائية متعددة المتغيرات التى سبق أن أشرنا إليها، يسمى التحليل العاملى Factor Analysis . ويهدف هذا الأسلوب لتحليل هذه العلاقات وذلك بتكوين مصفوفة لقيم معاملات الارتباط بين درجات الاختبار الجديد، ودرجات عدة اختبارات أخرى مناسبة، وإجراء تحليل إحصائى لهذه المصفوفة لتحديد أقل عدد من العوامل (التكوينات الفرضية) التى تسهم فى تفسير التباين فى قيم معاملات الارتباط داخل المصفوفة، وكذلك العوامل التى تسهم فى الأداء فى كل من هذه الاختبارات.

وبذلك نستطيع التوصل إلى تجمع من الاختبارات التى تقيس تكوينًا فرضيا معينًا، ودراسة محتوى هذه الاختبارات لتعرف طبيعة هذا التكوين الفرضى. ويمكن إجراء هذا التحليل باستخدام البرنامج الفرعى Factor ،وهو أحد برامج حزمة الحاسوب الإحصائية للعلوم الاجتماعية (SPSS).

ولكن نظراً لأنه من الضروري عدم الاقتصار على تعرف العلاقات الموجبة بين الاختبار الجديد الذى يفترض أنه يقيس تكويناً فرضياً معيناً وغيره من الاختبارات المشابهة التى تقيس التكوين الفرضى نفسه، وإنما أيضاً تعرف علاقته ببعض الاختبارات أو المقاييس التى لا تقيس هذا التكوين الفرضى، فقد اقترح كامبل Campbell وفيسك Fiske عام ١٩٥٩ أسلوباً للتحقق من ذلك اعتمدا فيه على مصفوفة أطلقا عليها «مصفوفة السمات المتعددة والطرق المتعددة Multitrait - Multimethod Matrix » يمكن باستخدامها الكشف عن علاقات الاختبار باختبارات مشابهة للتوصل إلى الصدق التقاربى للاختبار Convergent Validity، وعلاقته باختبارات مختلفة عنه للتوصل إلى الصدق التمايزى للاختبار Discriminent Validity.

وتعد مصفوفة السمات المتعددة والطرق المتعددة بمثابة تصميم تجريبى منظم يقدم أدلة عن كل من هذين النوعين من الصدق، ويتطلب إعداد هذه المصفوفة قياس عدد من التكوينات الفرضية بطرق متعددة، ولتوضيح هذا الأسلوب نفترض أننا قمنا بقياس ثلاث سمات مختلفة (الفهم اللفظى، والاستدلال العددى، والمعلومات العامة) بطرق ثلاث (تقديرات المعلمين، واختبارات ورقية، واختبارات فردية) وبذلك تشتمل المصفوفة على تسعة متغيرات، وندون فى خلايا هذه المصفوفة قيم معاملات الارتباط بين كل متغيرين منها، ويوضح جدول (٥ ـ ٣) هذه القيم الافتراضية:



		الطريقة الأولى			الطريقة الثانية			الطريقة الثالثة		
		أ١	ب١	جـ١	أ٢	ب٢	جـ٢	أ٣	ب٣	جـ٣
	أ١	٠,٩٣								
الطريقة الأولى	ب١	٠,٤٨	٠,٩٢							
	جـ١	٠,٢٢	٠,٤١	٠,٩٠						
	أ٢	٠,٥٧	٠,٢٠	٠,١٣	٠,٩٢					
الطريقة الثانية	ب٢	٠,٣٣	٠,٦٢	٠,٠٢	٠,٦٢	٠,٩٢				
	جـ٢	٠,١٨	٠,١٢	٠,٥٤	٠,٦١	٠,٤٨	٠,٨٩			
	أ٣	٠,٥٧	٠,٢٥	٠,١٠	٠,٧٠	٠,٣٨	٠,٣١	٠,٩٢		
الطريقة الثالثة	ب٣	٠,٢٠	٠,٦٢	٠,١٥	٠,٣٨	٠,٦٩	٠,٢٩	٠,٦٠	٠,٩٢	
	جـ٣	٠,١٢	٠,١٣	٠,٥٠	٠,٣٠	٠,٤٩	٠,٦٧	٠,٦٥	٠,٦١	٠,٨٨

جدول (٥ـ٣) يوضح مصفوفة السمات المتعددة والطرق المتعددة لمجموعة من البيانات الافتراضية

ويتضح من جدول (٥ ـ ٣) أن هنـاك قيما لمعاملات الارتبـاط تقع على القطر الرئيسي للمصفوفـة، وهذه تعد بمثابة قيم معامل الثبات؛ لأنها تشير إلى قياس السمة نفسها بالطريقة نفسها، أي(أ، أ، ، ب، ب، ، جـ، جـ، ، . . .) ويلاحظ أنها قيما مرتفعة.

وهناك قيم تقع داخل مثلثات تحـدها خطوط متقطعة، وهـذه تشير إلى قيم معامـلات الارتباط بين السـمات الثلاثة التي أجرى قياسـها بكل من الطـرق الثلاث، ومعظمها قيمًا منخفضة . كما أن هنـاك قيمًا تقع داخل مثلثات تحـدها خطوط متصلة (المثلثات الثلاثة التي تقع على الخط الرئيسي للمصفوفة) ، وهذه تشير إلى قيم معامل الارتباط عبر السـمات المقاسة بطريقة واحدة، وهي قيم مرتفعة نسبيًا. أما القيم التي تقع على القطرين اللذين يفصلان بين المثلثات التي تحدها خطوطًا متقطعة فإنها تشير إلى قيم معامل الارتباط بين السمة نفسها مقاسة بالطرق الثلاث، وهي قيم مرتفعة نسبيًا أيضًا. وهذه القيم تدل على معاملات الصدق ويقصد به الصدق التقاربي، وينبغي أن تكون هذه القيم موجبة ومرتفعة.

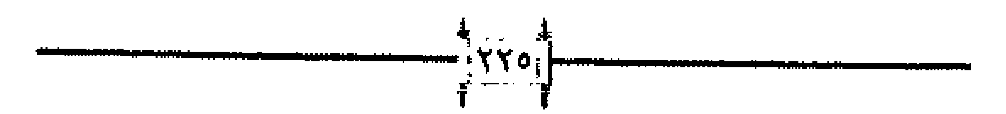
كما ينبغي أن تكون قيم معامل الارتباط بين الطرق المختلفـة التي تقيس السمة نفسها (معاملات الصدق) أكبر من قيم معامل الارتباط بين السمات المختلفة التي تقاس بالطريقة نفسها، وهي القيم المدونة داخل المثلثات التي تحدها خطوطًا متصلة.

فالتمييز بين السمات المتباينة أهم من التمييز بين طرق القياس المختلفة، كما أن قيم معاملات الصدق ينبغي أن تكون أكبـر من قيم معامل الارتباط بين كل من السمات المتبـاينة والطرق المختلـفة. وهي القيم المدونـة داخل المثلثات التي تحـدها خطوطًا متقطعة.

وهاتان الملاحظتان الأخيرتان تفيدان في تقديم أدلة عن الصدق التمايزي ، فما يهمنا هو معرفة ما إذا كان الارتباط بين السمـات كان نتيجة لاختلاف طريقة قياس هذه السمات مما قد يشير إلى تحيز طريقة القياس، أم أنه ارتباط فعلي. وهذا يعد من الأدلة المهمة التي تزودنا بها هذه المصفوفة في تقدير الصدق التقاربي والتمايزي التي ربما لا تزودنا بها الأساليب الأخرى، نظرا لاعتماد هذه الأساليب على طريقة قياس واحدة.

أساليب تعتمد على التجريب :

تعتمد هذه الأساليب على التدخل التجريبي لإحداث تغييرات في درجات الأفراد في اختبار ما كوسيلة لتعرف مدى تأثر الأداء بمعالجات أو متغيرات معينة مما يساعد في تأكيد بعض التفسيرات المتعلقة بنتائج الاختبار أو رفضها.



فإذا تأثرت درجات الاختبار بمؤثرات انتقالية، فإن هذا ربما يحد من تفسير الدرجات مما يضفي بعض الشك على صدق التكوين الفرضي للاختبار. فمثلاً إذا أردنا معرفة مدى تأثر اختبار يقيس فهم المقروء باستراتيجية تعامل الطالب مع المقروء، فإننا نطبق الاختبار على مجموعة من الطلاب بالطريقة المعتادة، ثم نطبق عليهم اختبارا مكافئًا بعد تدريبهم تدربًا سريعًا على استخدام استراتيجيات مختلفة لفهم المقروء، فإذا وجدنا تحسنا في درجاتهم، فإن هذا يكون دليلاً على أن الاختبار يتأثر بالاستراتيجية المستخدمة.

ويمكن أيضًا الاستعانة بمجموعة ضابطة يتم تدريبها بطريقة محدودة الفاعلية لمقارنة درجاتها بدرجات المجموعة التجريبية المشار إليها. كذلك يمكن التحقق مما إذا كان الاختبار تعتمد الإجابة عن مفرداته على الدافعية. فعندئذ يمكن تصميم تجربة يقوم فيها الفاحص بتعزيز مجهودات الطلاب وتبين أثر ذلك على درجاتهم. فالتدخل التجريبي في هذين المثالين يعد تدخلاً قصير الأمد؛ لأنه يحدث أثناء العملية الاختبارية أو قبلها بوقت قصير، ولكن يمكن أيضًا أن يستغرق التدخل التجريبي وقتًا أطول مثل أساليب العلاج النفسي أو الأساليب العلاجية التربوية في دراسات صدق التكوين الفرضي.

وكذلك للتحقق من صدق التكوين الفرضي لمقياس «القلق»، من المهم تحديد ما إذا كان المقياس يزودنا بنتائج تتفق مع نظرية معينة تتعلق بالقلق. فإذا أوضحت النظرية أن القلق يزداد في مواقف معينة تثير الإحباط فإنه يمكن إجراء تجربة لتحديد ما إذا كان الأفراد الذين يواجهون أحد هذه المواقف يحصلون على درجات مرتفعة في المقياس أكثر من الذين لم يواجهوا هذا الموقف.

كما يمكن أن تشمل الأساليب التي تعتمد على التجريب أدلة من مواقف طبيعية. فمثلاً إذا كان تعريف تكوين فرضي معين يتضمن أن هذا التكوين الفرضي لا يتأثر بالنضج أو المتغيرات البيئية فإنه يمكن التحقق من ذلك تجريبيا.

ولتوضيح ذلك نفترض أننا نود دراسة صدق التكوين الفرضي لاختبار جديد في الذكاء العام، فربما نفترض أن الذكاء العام يزداد بسرعة لدى الأطفال الصغار، بينما يزداد بدرجة أقل لدى المراهقين، ويقل ازدياده تدريجيا لحين بلوغ سن النضج.

لذلك نطبق الاختبار على عينات من الأفراد ممثلة لمجموعات عمرية مختلفة من سن ٣ أعوام إلى ٣٠ عامًا، ونرسم منحنيًا يمثل متوسطات درجات كل مجموعة عمرية، فإذا أكد المنحنى الفرض السابق، فإن هذا يقدم أحد الأدلة عن صدق التكوين



الفرضى للاختبار، آخذين فى الاعتبار أن الاختبار الذى يفترض أنه يقيس الذكاء ينبغى أن تعكس درجاته هذه النزعة النمائية.

أساليب تعتمد على التحليل المنطقى:

تعتمد هذه الأساليب على الفحص الدقيق للاختبار والأداء الذى يتطلبه، وإحداث تكامل بين نتائج هذا الفحص وبين النظرية التى يستند إليها الاختبار وآراء المُختبرين الذين سبق أن اختبروا باختبارات مشابهة، ويترتب على ذلك افتراض تفسيرات ربما تختلف عن التفسيرات التى اقترحها مُعد الاختبار.

ويرى كرونباك Cronbach أن هذا التحليل المنطقى من أهم مصادر التوصل إلى فروض بديلة فيما يتعلق بالأداء فى الاختبارات والمقاييس، فالمحكم الذى لديه خبرة سابقة كبيرة بالأخطاء التى شابت الاختبارات السابقة يمكنه أن يكتشف جوانب الضعف فى أداة القياس الجديدة.

ومع هذا فإننا لا نستطيع باستخدام هذا التحليل المنطقى للمحتوى أن ندحض صدق أداة القياس، وإنما نستطيع أن نقدم فروضاً بديلة يمكن التحقق من صحتها بالأساليب الإمبيريقية أو التجريبية.

كذلك من المفيد تحليل العمليات Process Analysis التى يستخدمها الأفراد فى أدائهم أو فى التوصل إلى إجاباتهم عن مفردات الاختبارات؛ لأن هذا التحليل يضيف أدلة جديدة تزيد من فهمنا للتكوينات الفرضية التى تقيسها هذه الاختبارات. إذ يمكن باستخدام نتائج ملاحظة هذه العمليات تقييم ما تسهم به من تباين فى درجات الاختبارات تقييماً أولياً.

فإذا افترضنا أن اختباراً معيناً يقيس كفاية الطلاب فى الرياضيات، ولكن بينت ملاحظة أداء مجموعة من الطلاب فى الاختبار أن غالبيتهم يخطئون فى قراءة الأسئلة، فعندئذ يصبح تفسير الدرجات المنخفضة موضع تساؤل. وعلى الرغم من صعوبة ملاحظة العمليات التى يستخدمها الأفراد فى إجابتهم عن مفردات الاختيار من متعدد، إلا أنه يمكن أن نطلب من كل منهم الإجابة بصوت مسموع، حيث يمكن أن يوضح الفرد جوانب الصعوبة ويقظته فى كشف أخطائه، كما يمكن أن يُطلب من الفرد بعد انتهائه من الإجابة أن يبرر اختياره للإجابة الصحيحة عن كل مفردة واستبعاده البدائل الأخرى للإجابات التى تشتمل عليها المفردة. وهذا ربما يلقى الضوء على المؤشرات التى أدت بالأفراد إلى الإجابة الصحيحة بعمليات غير مناسبة.

ونظراً لصعوبة إجراء هذا النوع من التحليل لما يتطلبه من وقت وجهد وتنظيم معين، فقد اتجه بعض الباحثين إلى بناء برامج للحاسوب تتضمن شبكة العمليات النموذجية المتعلقة بحل مشكلة أو إجراء مهمة معينة، ويتم إدخال استجابات الأفراد لمفردات الاختبار في الحاسوب، ويجرى البرنامج تصنيفًا لأنماط الأخطاء وترتيبها بحسب درجة أهميتها في حل المشكلة أو أداء المهمة.

والخلاصة أن صدق التكوين الفرضي يساعد في تحديد السمات التي يهدف الاختبار لقياسها، كما يفيد في بناء وتمحيص النظريات التربوية والسيكولوجية. وهذا يتطلب جمع أدلة ومعلومات متعددة لتلقى ضوءاً كافيًا على التكوينات الفرضية باستخدام الأساليب السابقة.

وقد بين ساكس Sax وزيلر Zeller أنه يمكن إيجاز عملية التوصل إلى أدلة تتعلق بصدق التكوين الفرضي في الخطوات التالية على الأقل:

(١) تبرير أهمية التكوين الفرضي من الناحية التربوية والنفسية على أن يكون التكوين الفرضي معرفًا تعريفًا إجرائيًا ويدل على سمة قابلة للقياس.

(٢) الاستناد إلى نظرية تربوية أو سيكولوجية أو اقتراح نموذج منطقي يوضح المفاهيم والعلاقات القائمة بينها.

(٣) التمييز بين التكوين الفرضي والتكوينات الفرضية الأخرى المماثلة لها.

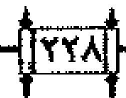
(٤) التوصل إلى أدلة من مصادر متعددة باستخدام الأساليب الارتباطية والتجريبية والمنطقية لتأكيد التكوين الافتراضي. وعندما تستخدم اختبارات متعددة في هذا الشأن فإنه يمكننا الحصول على الصدق التقاربي للتكوين الفرضي.

(٥) التوصل إلى أدلة نتأكد منها أن التكوين الفرضي لا يرتبط بعوامل وقتية أو دخيلة لكي نحصل على الصدق التمايزي للتكوين الفرضي.

(٦) إجراء تعديلات مستمرة في التكوين الفرضي بما يتفق والأدلة والمعلومات الجديدة المتجمعة.

منظور تكاملي لطرق تقدير الصدق والثبات:

اتضح لنا مما سبق أن صدق المحتوى والصدق المرتبط بمحك وصدق التكوين الفرضي ليست أنواعًا متباينة، وإنما جوانب متكاملة لعملية متسعة تتعلق بالتحقق من صدق تفسيرات واستخدامات الدرجات المستمدة من أدوات القياس التربوي والنفسي.



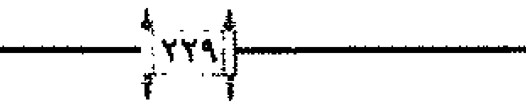
لذلك فإن الفصل بين هذه الجوانب ، أى دراسة أحدها بمعزل عن الجوانب الأخرى يؤدى إلى أدلة غير كافية حول صدق التفسيرات التى نتوصل إليها. لذلك ينبغى أن نضع فى اعتبارنا أن صدق أدوات القياس ليس صدقًا عامًا وإنما مرتبط بمواقف أو استخدامات معينة لمجموعات معينة من الأفراد.

وقد سبق أن لاحظنا من الشكل التخطيطى (٥ ـ ١) أن الأدلة التى تجمع للتحقق من صدق التكوين الفرضى تتضمن أدلة عن صدق محتوى أداة القياس وصدقها المرتبط بمحك، فصدق التكوين الفرضى إذن يعد بمثابة الإطار المتكامل الذى يجمع جوانب الصدق الثلاثة، ويمكن أن يجيب عن التساؤلات الثلاثة التى ذكرناها فى مستهل حديثنا عن مفهوم الصدق.

فتفسير صدق المحتوى يتعلق بالاستدلال على أداء الفرد فى النطاق الشامل للمفردات من أدائه فى اختبار يشتمل على عينة ممثلة من المفردات، ولكن إذا أردنا الاستدلال من ذلك على «قدرة الفرد» أو «معارفه» فإننا نكون بصدد الاستدلال على التكوين الفرضى الذى تنطوى عليه درجات الفرد. كذلك إذا أردنا التنبؤ بحصوله على تقدير مرتفع فى نهاية مدة دراسية معينة باستخدام درجاته فى عينة المفردات فإن تفسيرنا يتعلق بمحك الأداء المستقبلى.

فالاختبار عندئذ تتعدد جوانب الصدق فيه، ولا يقتصر صدقه على جانب واحد. كما أن الاختبار فى هذه الحالة يمكن استخدامه كأداة تفيد فى اتخاذ القرارات، وفى توضيح طبيعة السمة أو المتغير الذى يقيسه.

وهذا المنظور التكاملى للصدق يلعب دورًا رئيسًا فى عمليات بناء وتطوير أدوات القياس التربوى والنفسى. فكثير من الباحثين والممارسين يظنون أن الصدق يرتبط بالصيغة النهائية للاختبار. ولكن إذا أمعنا الفكر قليلاً نجد أن صدق المحتوى يتعلق بالمرحلة الأولى فى بناء الاختبار، وهى تجميع عينة ممثلة من المفردات للنطاق السلوكى الشامل المراد قياسه، ويمكن أيضًا انتقاء هذه العينة بأساليب الصدق المرتبط بمحك من حيث تمييزها بين الناجحين والراسبين، أو بين المتمكنين وغير المتمكنين فى ضوء محك أداء معين. ويمكن حذف المفردات التى لا تحقق متطلبات تكوين فرضى معين تقيسه المفردات، وربما نجرى ذلك بالتحقق من الاتساق الداخلى أو تجانس عينة المفردات. وينطبق ذلك بوضوح على اختبارات الذكاء والاستعدادات والاختبارات التحصيلية، وربما ينطبق أيضًا على بعض مقاييس الجوانب الوجدانية والشخصية.



ولذلك فإن مراحل بناء وتطوير أدوات القياس تتفق بل وتتكامل مع أساليب وإجراءات تقدير الصدق بجوانبه الثلاثة.

كذلك يرتبط مفهوم الصدق بمفهوم الثبات ارتباطًا وثيقًا، فقد سبق أن أوضحنا أن الاختبار لا يمكن أن يتسم بالصدق في استخدامات معينة دون أن تكون درجاته متسقة، كما أن قيمة مؤشر الثبات تعد بمثابة الحد الأعلى لقيمة معامل صدق الاختبار. وتؤثر قيم معامل الثبات في درجة العلاقة بين الاختبار التنبؤى واختبار المحك. وكذلك تسهم قيمة معامل الاستقرار لدرجات اختبار ما في زيادة ثقتنا بصدق التكوين الفرضي للاختبار؛ نظراً لأننا نهتم بالسمات التي تتميز بالثبات النسبي.

كما أن معامل الاتساق الداخلي أو معامل التجانس للاختبار يقدم دليلا آخر على صدق التكوين الفرضي.

لذلك يرى كرونباك Cronbach وجليزر Glaser أن مفهوم الصدق والثبات مفهومان مترابطان ويمكن أن يندرجا تحت اسم «مقاييس قابلية التعميم Measures of Generalizability»،فالفرق الرئيسي بينهما يكمن في الأبعاد التي نود التعميم عليها. لذلك تعد نظرية كرونباك المتعلقة بإمكانية التعميم Generalizability Theory التي أشرنا إليها في مناقشتنا لمفهوم الثبات من التطورات المعاصرة المهمة التي أسهمت في إبراز التكامل بين مفهومي الثبات والصدق.

ولذلك سوف نوضح هذه النظرية بمزيد من التفصيل كما سبق أن أشرنا في الفصل الرابع عشر.

خلاصة

يُعد الصدق من الخصائص السيكومترية الأساسية للاختبارات والمقاييس، ونعنى بالصدق مدى فائدة القياسات فى اتخاذ قرارات موائمة لغرض معين. والاختبار لا يتسم بالصدق دون أن تكون درجاته متسقة، وهذا يعنى أن مفهومى الصدق والثبات مترابطان. ونظراً لأن الاختبار يمكن أن يستخدم فى أغراض متعددة، فإنه يمكن أن يكون للاختبار عدة مؤشرات للصدق وفقًا لأغراض استخداماته، ونوع القرارات التى يمكن اتخاذها استناداً إلى درجاته.

لذلك فإن هناك جوانب متعددة للصدق تناظر ثلاثة أقسام رئيسة من القرارات، أحدها: صدق المحتوى، ويتعلق بمدى تمثيل محتوى الاختبار للنطاق السلوكى الشامل للسمة المراد الاستدلال عليها، والثانى الصدق المرتبط بمحك، ويتعلق بمدى صلاحية الاختبار أو المقياس فى أغراض التنبؤ من أجل اتخاذ قرارات عملية مثل قرارات الانتقاء أو التصنيف. ويمكن تمييز نوعين من هذا الصدق أحدهما: الصدق التنبؤى الذى يدل على مدى ارتباط أداء الأفراد فى الاختبار بأدائهم فى محك مستقبلى، والآخر: الصدق التلازمى، ويدل على اقتران تباين درجات اختبار بدرجات اختبار أو مقياس آخر يُطبق فى الوقت نفسه تقريبًا على مجموعة من الأفراد. وتتعدد طرق تقدير الصدق التنبؤى، من أهمها إيجاد الارتباط بين الاختبار التنبؤى واختبار المحك، أو استخدام معادلة الانحدار فى التنبؤ بدرجات الاختبار المحك أو استخدام جداول التوقُّع ، أو الاستناد إلى نظرية المنفعة. غير أن هناك بعض المشكلات المتعلقة بالمحك، وما ينبغى أن تتوافر فيه من شروط، مثل ثبات درجاته، وموضوعيته، ومواءمته لنوع القرار المراد اتخاذه . كما أن الصدق المرتبط بمحك يتأثر بمجموعة من العوامل من أهمها: مدى تجانس عينة الأفراد المختبرين وعددها، وثبات درجات المحك وعدد مفرداته، ومدى تأثره بمتغيرات دخيلة، والمدة الزمنية التى تفصل بين تطبيق الاختبار التنبؤى واختبار المحك.

أما الجانب الثالث من الصدق فيتعلق بصدق التكوين الفرضى الذى يتناول العلاقة بين نتائج الاختبار أو المقياس وبين المفهوم النظرى الذى يهدف الاختبار لقياسه. ونظراً لأن التكوينات الفرضية هى التى يتركز عليها الاهتمام فى القياس، فإن هذا الجانب من الصدق قد نال قدراً كبيراً من الجهود البحثية، وتعددت منهجيات وأساليب التحقق منه. فهناك أساليب تعتمد على الارتباطات، وأساليب تعتمد على التجريب، وأساليب تعتمد على التحليل المنطقى والفحص الدقيق للاختبار والأداء الذى يتطلبه فى ضوء النظرية التى يستند إليها الاختبار.

غير أنه لا يجب النظر إلى الجوانب الثلاثة للصدق على أنها جوانب متباينة، وإنما تعد جوانب متكاملة لعملية متسعة تتعلق بالتحقق من صدق تفسيرات واستخدامات الدرجات المستمدة من أدوات القياس. كما أن مفهومى الصدق والثبات تُعدان من المفاهيم المترابطة، ويندرجا تحت مقاييس «قابلية التعميم» التى سنوضحها فى الفصل الرابع عشر.

الفصل السادس

معايير الاختبارات ومحكات الأداء

* مقدمة
* معايير الاختبارات مرجعية الجماعة أو المعيار
* أنواع الجماعات المرجعية
* خصائص عامة للجماعة المرجعية
* أنواع معايير الاختبارات
* أولا : الدرجات المحوّلة
* الدرجات التائية الخطية والمعيارية
* التساعيات المعيارية
* ثانيًا : معايير المئينيات والرتب المئينية
* ثالثًا : معايير الارتقاء
* معايير العمر الزمنى
* معايير الفرق الدراسية
* رابعًا : الصفحة النفسية أو التربوية (البروفيل النفسى أو التربوى)
* محكات الأداء فى الاختبارات مرجعية المحك

مقدمة :

يعد مفهوم معايير الاختبارات Tests Norms من المفاهيم الأساسية المتعلقة بتفسير درجات الاختبارات مرجعية الجماعة أو المعيار Norm - Referenced Tests ، فالدرجة التى يحصل عليها فرد فى اختبار ما والتى تسمى الدرجة الخام Raw score لا يكون لها معنى ويصعب تفسيرها ما لم يتم إسنادها إلى نظام مرجعى Reference system ، فهذا النظام هو الذى يسمح باستخلاص معلومات مفيدة من درجات الاختبار.

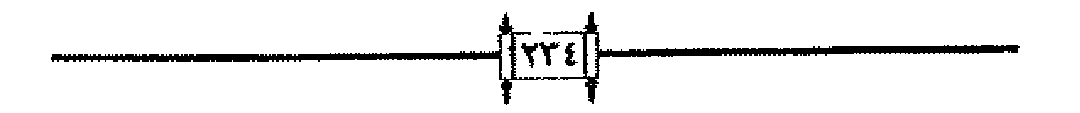
والنظام المرجعى المناسب لهذا النوع من الاختبارات يعتمد على استخدام معلومات يتم الحصول عليها من جماعة معينة تسمى الجماعة المرجعية Norm Group ، وهذه الجماعة تكون خصائصها محددة ومعلومة لمستخدمى الاختبار، لذلك تسمى الاختبارات التى تستخدم هذا النظام الاختبارات مرجعية الجماعة أو المعيار.

وهناك نظام آخر لا يعتمد على جماعة مرجعية، وإنما يعتمد على محك أداء متوقَّع Criterion ، ويتحدد هذا المحك فى ضوء الكفايات التى تشتمل على معارف ومهارات متوقعة ينبغى أن يمتلكها الفرد.

ونظراً لأن تفسير درجات الاختبارات فى هذه الحالة يستند إلى هذا المحك فإن الاختبارات عندئذ تسمى الاختبارات مرجعية المحك Criterion- Referenced Tests . وقد أشرنا إلى التمييز بين هذين النوعين من الاختبارات فى الفصل الأول، وسوف نتناولهما بالتفصيل فى الفصل الثامن.

وينصب الاهتمام فى النظام الأول على مقارنة الدرجة التى يحصل عليها فرد فى اختبار مرجعى الجماعة بأداء أقرانه بهدف ترتيب درجات الأفراد فى الاختبار بالنسبة لبعضهم البعض، وبذلك يمكن تحديد المركز النسبى للفرد بين أقرانه فى ضوء معيار جماعته المرجعية.

إذ لا نستطيع إجراء هذه المقارنة باستخدام الدرجات الخام أو النسب المئوية من العدد الكلى لأسئلة الاختبار التى يجيب عنها كل فرد إجابة صحيحة. فلكى نحدد على وجه الدقة مركز الفرد بالنسبة للجماعة المرجعية ينبغى تحويل هذه الدرجات الخام إلى نوع آخر من الدرجات يسمح بهذه المقارنة، مثل الدرجات المعيارية، والدرجات التائية، والمئينيات، والرتب المئينية، وغيرها من الدرجات المحوَّلة التى ناقشناها بإسهاب فى الفصل الثالث.



ولتوضيح ذلك نفرض أن درجة فرد في اختبار للذكاء ١٣٢، فهنا ربما يتبادر إلى الأذهان أن هذا الفرد أجاب عن ١٣٢ مفردة إجابة صحيحة، أو أن الفرد حصل على ١٣٢ نقطة من المجموع الكلي للنقاط وهو ٢٠٠ مثلاً.

ولكن هذه الدرجة الخام لا تنقل لنا صورة واضحة عن مركزه بالنسبة لأقرانه في الذكاء، لذلك لابد من إجراء نوع من التحويل على هذه الدرجة لكي يتسنى لنا مقارنتها بغيرها من الدرجات.

ومن هنا تتضح أهمية معايير الاختبارات، فدرجات اختبارات الذكاء والشخصية والاختبارات التحصيلية المقننة وغيرها من الاختبارات والمقاييس تفسر عادة باستخدام جداول المعايير.

- معايير الاختبارات مرجعية الجماعة أو المعيار:

إن المعايير عبارة عن مجموعة من الدرجات المشتقة بطرق إحصائية معينة من الدرجات الخام كما ذكرنا بحيث تأخذ بعين الاعتبار توزيع الدرجات الخام المستمدة من تطبيق الاختبار على عينة عشوائية ممثلة للمجتمع المستهدف.

وقبل أن نوضح أنواع المعايير التي تستخدم في تفسير الدرجات ينبغي أن نلقي مزيداً من الضوء على أنواع الجماعات المرجعية التي يستند إليها في التوصل إلى هذه المعايير، وخصائصها وكيفية انتقائها.

أنواع الجماعات المرجعية:

تعتمد معايير الاختبارات مرجعية الجماعة أو المعيار على الجماعة المرجعية التي تستمد منها هذه المعايير Norm Group . لذلك ينبغي أن تكون خصائص هذه الجماعة محددة تحديداً جيداً ليتسنى تفسير درجات هذه الاختبارات تفسيراً دقيقاً.

ولتوضيح ذلك نفترض أننا قارنا درجة طالب في اختبار تحصيلي في الفيزياء بدرجات مجموعة من الطلاب المبتدئين في دراسة هذه المادة في المرحلة الثانوية، فإن ترتيب هذا الطالب سيكون مرتفعاً بالنسبة لهذه المجموعة، أما إذا قارنا درجته بدرجات مجموعة أخرى من الطلاب المتخصصين في الفيزياء، فإن ترتيبه سيكون منخفضاً. أى أننا إذا لم نعرف شيئاً عن طبيعة الجماعة التي نقارن الطالب بها، فإنه لا يمكن التوصل إلى استنتاج سليم فيما يتعلق بالمركز النسبي للطالب في الفيزياء. ومن هنا تبدو أهمية تحديد خصائص الجماعة المرجعية التي سوف يقارن بها الطلاب.

فخصائص الجماعة المرجعية ينبغى أن تكون مشابهة لخصائص أفراد الجماعة الذين سيطبق عليهم الاختبار فيما بعد من أجل تقويم السمة التى يقيسها لديهم فى ضوء مركزهم النسبى.

لذلك يبرز التساؤل التالى: ما الخصائص التى ينبغى أن تتوافر فى الجماعة المرجعية؟ والحقيقة أن هذه الخصائص تتباين بحسب طبيعة المعايير المستخدمة ، فالمعايير الوطنية National Norms تختلف عن المعايير المحلية Local Norms وكلاهما يختلف عن المعايير المتعلقة بموقع واحد مثل مؤسسة أو مدرسة معينة Institution Average Norms ،وكذلك معايير فئات خاصة - Special Norm Groups. لذلك ربما يكون من المناسب توضيح كل من هذه المعايير فيما يلى:

المعايير الوطنية: تفيد هذه المعايير فى تقويم البرامج التعليمية باستخدام الاختبارات التحصيلية المقننة؛ وذلك للتحقق مما إذا كان أداء طلاب نهاية مرحلة تعليمية معينة أو فرقة دراسية أعلى أو أدنى من المتوسط على المستوى الوطنى.

لذلك تقوم مؤسسات متخصصة بإعداد هذه الاختبارات نظراً لما تتطلبه عمليات بنائها وإعداد معاييرها الوطنية من وقت وجهد كبير وكلفة مرتفعة. فالجماعة المرجعية الوطنية National Norms Group يفترض أنها بمثابة عينة ممثلة لطلاب الوطن فى مرحلة تعليمية معينة، غير أن تحديد هذه العينة وتمثيلها للوطن يختلف باختلاف وجهة نظر المؤسسة المعينة.

لذلك يصعب مقارنة المعايير الوطنية للاختبارات التى تقوم مؤسسات مختلفة بإعدادها. غير أن انتقاء هذه العينة ينبغى أن يأخذ بعين الاعتبار الخصائص الديموجرافية المختلفة، مثل المناطق الجغرافية، والفرق الدراسية، وحجم المدرسة وموقعها، والحالة الاجتماعية والاقتصادية، والنوع والتركيبة الطلابية، وغير ذلك من الخصائص ذات الأهمية.

ويتم تحديد ذلك بالتفصيل فى دليل الاختبار لكى يتحقق مستخدمو الاختبار من مدى تمثيل العينة الوطنية للمجتمع المستهدف. وأحياناً تعد المؤسسة المعنية معايير مستقلة للمجموعات الفرعية فى كل من هذه الخصائص على حدة إذا تبين أن كلا من هذه المجموعات تختلف عن الأخرى فى أى من هذه الخصائص.

ويتم اختيار هذه العينة باستخدام أساليب المعاينات العشوائية Random Sampling أو المعاينات العشوائية الطبقية التى تراعى الخصائص أو المتغيرات السابقة Startified Random Sampling . ويجب أن تكون العينة المنتقاة كبيرة بدرجة كافية لضمان ملاءمة المعايير للمجتمع المستهدف، وتقليل أخطاء المعاينات Sampling Errors .

وتستخدم المعايير الوطنية أيضًا في اختبارات الاستعدادات، مثل اختبار الاستعداد الأكاديمي Scholastic Aptitude Test (SAT) الذي يستخدم في الولايات المتحدة الأمريكية للتنبؤ بالنجاح الأكاديمي، حيث اعتمدت معاييره على عينات من طلاب نهاية المرحلة الثانوية اشتملت على عدة آلاف من الطلاب من مختلف الولايات لكل اختبار فرعي.

غير أننا نفتقر في الوطن العربي إلى مثل هذه الاختبارات التي تعتمد على معايير وطنية أو إقليمية يمكن باستخدامها مقارنة أداء الطالب في مجال دراسي معين بمتوسط أداء أقرانه على هذين المستويين.

المعايير المحلية: تعد هذه المعايير في بعض الأحيان أكثر فائدة من المعايير الوطنية. فنادراً ما نجد تطابقًا بين خصائص الجماعة المرجعية التي تستمد منها المعايير الوطنية وخصائص الجماعة التي يطبق عليها الاختبار فيما بعد، إذ إن هذه المعايير يختلف معناها بمرور الزمن وتغير خصائص الأفراد مما يؤدي إلى صعوبات في تفسير الدرجات تفسيراً ذي معنى.

فمثلاً التركيبة الطلابية لمجتمع محلي معين تختلف عادة عن نظيرتها في الجماعة المرجعية، لذلك ربما تفضل المؤسسات التعليمية جمع بيانات من طلاب هذا المجتمع وإعداد معايير محلية استناداً إلى هذه البيانات وذلك لتفسير درجاتهم في الاختبارات مرجعية الجماعة أو المعيار تفسيراً أكثر مواءمة لظروف المجتمع المحلي الذي تتواجد فيه المؤسسة التعليمية. أي أن المعايير المحلية تسمح بمقارنة أداء الطالب بأقرانه في المنطقة التي يعيش فيها، ولا تسمح بالمقارنة خارج هذه المنطقة.

وكثير من اختبارات الذكاء مثل اختبار ويكسلر Wechsler Test يتضمن معايير تستند إلى الراشدين بعامة مما يؤدي إلى صعوبات في تفسير درجات بعض الأفراد. فإذا كانت درجة طالب في الاختبار تدل على أن ذكاءه أعلى من متوسط أقرانه الذين في مستواه العمري اعتماداً على معايير الكبار بعامة، فإنه يمكن أن يكون ذكاؤه أقل من متوسط الطلاب الملتحقين بالكليات. فالتنبؤ بنجاحه عندئذ في إحدى الكليات يتطلب مقارنة درجته بمتوسط أداء هؤلاء الطلاب بدلا من مقارنتها بمتوسط أداء الراشدين بعامة، وهذا يتطلب معايير مشتقة من درجات هؤلاء الطلاب في الاختبار.

وإعداد المعايير المحلية لا يحتاج إلى جهود مؤسسات متخصصة وإنما ربما يحتاج إلى استشارتها فقط، حيث إن هذه المعايير تعتمد على جماعة مرجعية محلية مما ييسر عملية انتقاء العينة واختبارها وإعداد المعايير باستخدام الحاسوب .

كما أنها تيسر تفسير درجات الاختبار للمعلم والطالب وولى أمره والمجتمع المحلى، وتفيد المعلم فى تتبع التقدم الدراسى لتلاميذه ومقارنة أدائه فى المجالات الدراسية المختلفة بالنسبة لأقرانه.

ومن الجدير بالذكر أنه يمكن استخدام المعايير الوطنية والمعايير المحلية معًا حيث إنهما متكاملان. فإذا أردنا مقارنة تقدم مجموعات من الطلاب فى برامج تعليمية متماثلة، فإن المعايير المحلية تكون مناسبة، ولكن إذا أردنا أيضًا مقارنة الأداء المحلى بالأداء على المستوى الوطنى، فإن المعايير الوطنية تصبح ضرورية ومكملة للمعايير المحلية.

معايير مؤسسة أو مدرسة معينة: تفيد هذه المعايير فى مقارنة متوسط أداء طلاب فرقة دراسية معينة فى إحدى المدارس بمتوسط أداء الفرقة المناظرة فى مدارس أخرى. ونلاحظ أن الاهتمام هنا لا يتركز على مقارنة طالب معين بأقرانه وإنما مقارنة متوسط أداء الفرقة بعامة.

لذلك ينبغى تحديد متوسطات أداء هذه الفرقة لعينة ممثلة من المدارس على المستوى الوطنى، وترتيب هذه المتوسطات ترتيبًا نسبيًا. وعادة يكون تباين توزيع هذه المتوسطات أقل من تباين توزيع متوسط درجات الطلاب.

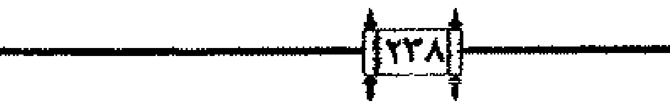
كذلك يمكن أن تعد مؤسسة معينة معايير خاصة للاختبارات التى تستخدم فى انتقاء المتقدمين لشغل وظائف بها، أو يقوم مكتب القبول بالجامعات بإعداد معايير لاختبارات الاستعدادات التى يطبقها على الطلاب المتقدمين للالتحاق بكلية معينة.

معايير فئات خاصة: تحتاج بعض الاختبارات إلى معايير لفئات خاصة من ذوى الإعاقات العقلية أو الجسمية مثل اختبارات الذكاء، أو لفئة معينة من الأفراد مثل طلاب الفرقة الأولى بكلية الطب، أو الطلاب المتقدمين للالتحاق بكليات الهندسة، أو أفراد بعض المهن، وكذلك لمؤسسات معينة مثل المدارس الخاصة، أو المعاهد الدينية، أو الشركات الاستثمارية، وهكذا. ففى مثل هذه الحالات ربما تبدو ضرورة إعداد معايير خاصة لهذه الفئات للاسترشاد بها فى اتخاذ قرارات مناسبة بشأنهم.

كما يمكن إعداد معايير خاصة لاختبارات الاستعدادات الموسيقية، والجمالية، والكتابية، والميكانيكية، وبعض الاختبارات التحصيلية المقننة، مثل اختبارات اللغات الأجنبية، والحاسبات الآلية، وما شابه ذلك.

خصائص عامة للجماعة المرجعية :

على الرغم من اختلاف الجماعات المرجعية تبعًا لطبيعة هذه الجماعات ومدى اتساع عيناتها، إلا أنه ينبغى أن تتوافر فيها بعض الخصائص العامة لكى تكون هذه



الجماعات مناسبة للمعايير التي تستمد منها. ومن أهم هذه الخصائص التى حددتها الجمعية الأمريكية لعلم النفس (APA) ما يلى:

تمثيل عينة الأفراد للمجتمع المستهدف: عند انتقاء عينة من الأفراد من أجل بناء المعايير ينبغى مراعاة أن تكون هذه العينة ممثلة تمثيلا كافيًا للمجتمع المستهدف.

فإذا اقتصرت العينة على إحدى المدارس أو المناطق الجغرافية، فإن العينة فى هذه الحالة تكون متحيزة. إذ ربما تكون هذه المدرسة متميزة أو أداء طلابها أقل من متوسط أداء طلاب مجتمع هذه المدرسة أو هذه المنطقة، فنوع المعايير سواء أكانت وطنية أم محلية أم خاصة يفرض طبيعة العينة المستخدمة. فإذا كانت المعايير وطنية، فإنه ينبغى التأكد من تمثيلها تمثيلاً متوازنًا للأفراد من مختلف المناطق، ومتناسبًا مع المقيمين فى الريف والحضر ، والذكور والإناث ، وغير ذلك من المتغيرات الديموجرافية.

وعلى الرغم من أن العينة العشوائية ربما تناسب معايير الاختبارات التى تقتصر على مدينة أو منطقة واحدة، حيث يمكن انتقاء عدة مئات من الأفراد عشوائيًا، إلا أنها تصبح غير ممكنة إذا كانت المعايير وطنية. فانتقاء عينة عشوائية كبيرة الحجم من مختلف محافظات الوطن يتطلب جهودًا كبيرة ونفقات باهظة، بل وربما تؤدى إلى زيادة أخطاء المعاينات.

لذلك يفضل استخدام العينات العشوائية الطبقية حيث يتم تصنيف المجتمع المستهدف وفقًا لعدد من المتغيرات التى أشرنا إليها وانتقاء نسبة معينة من كل مجموعة تصنيفية بحسب عددها فى المجتمع، وبذلك تقل أخطاء المعاينات ويقتصد الجهد والكلفة. غير أنه ينبغى مراعاة أن تمثيل العينة للمجتمع لا يعنى زيادة عددها فحسب، وإنما يعنى أيضًا مراعاة عدم تحيزها استنادًا إلي تصميم معاينات Sampling Design جيد. فحجم العينة غير المتحيزة إذا كان مناسبًا تكون المعايير المستمدة منها متميزة بالاستقرار النسبى بحيث إذا استخدمت عينة أخرى مشابهة نحصل على نتائج مماثلة ، وهذا ما يجب أن يتوافر فى معايير الاختبارات التربوية والنفسية. وربما كان هذا هو أحد الأسباب التى تدعو إلى إعداد معايير محلية إلى جانب المعايير الوطنية.

حداثة المعايير: إن معايير الاختبارات تتغير بتغير خصائص الأفراد بمرور الزمن وبخاصة إذا كانت الاختبارات تقيس خصائص تتأثر تأثرًا ملحوظًا بما يطرأ على البيئة من تطورات. فمثلاً المفردات اللغوية تنمو وتتغير دلالتها، والمناهج الدراسية تتطور، والبيئة التعليمية تحدث فيها تعديلات مستمرة. لذلك فإن معايير الاختبارات التحصيلية مثلاً تتأثر بجميع هذه العوامل التى تنعكس على الطلاب، بل وربما تتأثر اختبارات الاستعدادات ومقاييس الشخصية أيضًا ببعض هذه العوامل وغيرها.



وأحيانًا يظل محتوى الاختبارات كما هو على الرغم من التغيرات التي ربما تحدث للأفراد نتيجة التطورات المستمرة التي تحدث في المجتمع مما يجعل معايير هذه الاختبارات غير صالحة. فالجماعة المرجعية التي استمدت منها معايير اختبار معين منذ عدة سنوات لا تصبح مناسبة للاختبار في الوقت الحاضر.

لذلك ينبغي تحديث معايير الاختبارات بصفة مستمرة وفقًا لما يطرأ على الخصائص المحددة للجماعة المرجعية من تغير، أو ما يطرأ على محتوى الاختبار من تطوير. ولا ينبغي التحقق من حداثة المعايير بالنظر إلى تاريخ نشر الاختبار، إذ ربما تصدر طبعات جديدة من الاختبار نفسه دون إعادة النظر في الجماعة المرجعية التي استمدت منها معايير الاختبار.

مواءمة المعايير لمن سيطبق عليهم الاختبار: فالجماعة المرجعية تكون موائمة Relevant إذا اتفقت خصائصها مع خصائص من سيطبق عليهم الاختبار. فإذا كانت هذه الجماعة تمثل أفراد مهنة معينة، فإنه ينبغي أن تكون خصائصها مماثلة لخصائص أفراد المهنة نفسها الذين نود اختبارهم. ولكن إذا استُخدم الاختبار في أغراض متعددة، فإنه ينبغي أن تُستمد معاييره من أكثر من جماعة مرجعية واحدة مناسبة.

أنواع معايير الاختبارات:

تتعدد أنواع معايير الاختبارات التي تُستخدم في مقارنة درجة الفرد في اختبار ما بدرجات جماعة مرجعية من أقرانه. وتُشتق جميع هذه المعايير من الدرجات الخام للاختبار. غير أننا سوف نركز في عرضنا على المعايير شائعة الاستخدام في الاختبارات والمقاييس المنشورة لكي يتمكن الباحث ومستخدم الاختبار من تفسير الدرجات المستمدة منها في ضوء المعيار المناسب.

ويمكن تصنيف هذه الأنواع من المعايير بحسب طريقة تفسير كل منها للمركز النسبي للفرد فيما يقيسه الاختبار أو المقياس، ويوضح الشكل التخطيطي (٦ ـ ١) التالي هذا التصنيف المقترح:

معايير الاختبارات

معايير اداء نسبى (اختبارات مرجعية الجماعة)

محكات اداء مطلقة (اختبارات مرجعية المحك)

نسب مئوية تحدد مستويات الاداء

معايير الارتقاء

مئينيات ورتب مئينية

درجات محولة

معايير مكافئ العمر الزمنى

معايير مكافئ الفرق الدراسية

درجات معيارية خطية

درجات تائية خطية

تساعيات معيارية

نسبة الذكاء الانحرافية

درجات معيارية اعتدالية

درجات تائية اعتدالية

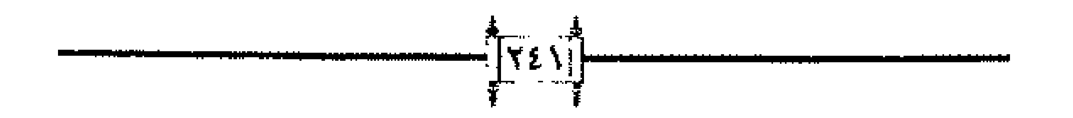
شكل (٦ ـ ١) يوضح أنواع معايير الاختبارات مرجعية الجماعة

ومحكات الاداء فى الاختبارات مرجعية المحك

(من تصميم المؤلف)

أولا : الدرجات المحولة :

تستخدم كثير من الاختبارات المنشورة فى الآونة الأخيرة معايير تعتمد على الدرجات المعيارية المحوّلة مثل الدرجات المعيارية الاعتدالية Normalized Z Scores والدرجات التائية الخطية Linear T Scores ، والدرجات التائية المعيارية Normalized



T Scores، والتساعيات المعيارية Stanines، ودرجات معيارية محوَّلة خاصة ببعض اختبارات القبول بالكليات فى الولايات المتحدة الأمريكية ، مثل اختبار الاستعداد الأكاديمى Scholastic Aptitude Test (SAT)، واختبار سجل الدراسات العليا Graduate Record Examintion (GRE) . وقد أعد هذان الاختباران مكتب امتحانات القبول بالكليات College Entrance Examination Board (CEEB) . كما أن معايير بعض اختبارات الذكاء مثل اختبار ويكسلر Wechsler Intelligence Test تعتمد على الدرجات المحولة مثل نسبة الذكاء الانحرافية، Deviation IQ (DIQ)، واختبار التصنيف العام لأفراد الخدمة العسكرية Army General Classification Test يعتمد على معايير يشار إليها بالحروف الأولى لاسم الاختبار (AGCT) . كما تستخدم بطاريات الاختبارات الصفحات التربوية أو النفسية Profiles . وبعض هذه الاختبارات تم تعريبه وتقنينه على البيئة المصرية وبعض البيئات العربية الأخرى مما سيتضح عند مناقشة الاختبارات والمقاييس التربوية والنفسية فى فصول لاحقة . ويمكن أن يرجع القارئ إلى الفصل الثالث للاطلاع على الدرجات المحوَّلة بأنواعها المختلفة وكيفية حسابها. وسوف نقتصر هنا على توضيح أهم هذه الدرجات المحوَّلة التى تُستخدم كمعايير للاختبارات والمقاييس.

الدرجات التائية الخطية والمعيارية :

سبق أن ذكرنا فى الفصل الثالث أن الدرجات التائية الخطية نحصل عليها بإجراء تحويل معين على الدرجات المعيارية الخطية، وذلك بضرب الدرجة المعيارية فى ١٠ وجمع ٥٠ على حاصل الضرب، وبذلك يصبح متوسط الدرجة التائية الخطية ٥٠، وانحرافها المعيارى ١٠.

وإذا جعلنا شكل توزيع الدرجات الخام اعتداليًّا يصبح توزيع الدرجات المعيارية الخطية والدرجات التائية الخطية اعتداليًّا ،بغض النظر عن شكل توزيع الدرجات الخام قبل جعله اعتداليًّا، وعندئذ نحصل على الدرجات المعيارية الاعتدالية، والدرجات التائية الاعتدالية. وتدون فى أدلة الاختبارات الدرجات الخام وما يقابلها من درجات تائية خطية أو اعتدالية، فإذا كان متوسط درجات اختبار ما ١٠٠ وانحرافه المعيارى ١٦ وحصل فرد على الدرجة ١١٦ فى الاختبار، فإن الدرجة المعيارية الخطية المقابلة لدرجته الخام = $\frac{١١٦ - ١٠٠}{١٦}$ = + ١ ، وهذه بدورها تقابل الدرجة التائية الخطية ١٠(+١) + ٥٠ = ٦٠ . وهذا يدل على أن درجته فى الاختبار تفوق المتوسط بقدر وحدة انحراف معيارى واحد (١٠ وحدات تائية) بغض النظر عن شكل توزيع الدرجات الخام فى الاختبار.

أما إذا جعلنا شكل توزيع الدرجات الخام اعتداليًا، فإنه يمكن تفسير درجة الفرد في ضوء معيار المئينيات الذى ناقشناه فى الفصل الثالث، وسوف نوضح استخداماته فى تفسير درجات الاختبارات بعد قليل.

معايير اختبار الاستعداد الأكاديمي :

أعد هذا الاختبار مكتب امتحانات القبول بالكليات في الولايات المتحدة الأمريكية (CEEB) عام ١٩٤١ ،ويطبق عادة لانتقاء الطلاب المتقدمين للالتحاق بالكليات. ويشتمل الاختبار على اختبارين فرعيين أحدهما لفظى والآخر رياضى. وطبق على مجموعة كبيرة من الطلاب بلغت ١٠٦٥٤ طالبًا. وتعتمد معايير كل من هذين الاختبارين الفرعيين على تحويل الدرجات الخام إلى درجات معيارية معدلة (CEEB) متوسطها ٥٠٠ وانحرافها المعيارى ١٠٠ .ويجرى ذلك بضرب الدرجات المعيارية الخطية (Z) المقابلة للدرجات الخام فى ١٠٠ وإضافة ٥٠٠ إلى حاصل الضرب، وبذلك يتحول توزيع الدرجات المعيارية الخطية إلى توزيع جديد قيمة كل من متوسطه وانحرافه المعيارى أكبر من قيمتهما فى معيار الدرجة التائية T Score.

وقد أجريت بعض التعديلات الإحصائية على درجات الطلاب الذين اختبروا بعد عام ١٩٤١ لضمان توحيد تفسير الدرجات فى ضوء معايير الاختبار من عام إلى آخر.

معايير اختبار التصنيف العام لأفراد الخدمة العسكرية، واختبار ويكسلر للذكاء :

فالاختبار الأول كان يستخدم لتصنيف الأفراد الملتحقين بالجيش فى الولايات المتحدة الأمريكية وتسكينهم فى الوحدات المختلفة، ويعد من الاختبارات شائعة الاستخدام فى هذا الغرض. واعتمدت معاييره (AGCT) على تحويل الدرجات الخام إلى درجات معيارية معدلة متوسطها ١٠٠ وانحرافها المعيارى ٢٠، وهما قيمتان اختياريتان.

أما اختبار ويكسلر للذكاء فيعد من الاختبارات المعروفة، واعتمدت معايير درجات اختباراته الفرعية على تحويل الدرجات الخام إلى درجات معيارية معدلة متوسطها ١٠ وانحرافها المعيارى ٣، بينما اعتمدت معايير درجاته الكلية على تحويل الدرجات المعيارية المقابلة لها إلى درجات معيارية معدلة متوسطها ١٠٠ وانحرافها

المعـيارى ١٥ ، وهما قـيمتان اخـتياريتـان أيضا. وتـسـمى هذه المعايـير نسبـة الذكاء الانحرافية (DIQ) Deviation IQ .

فإذا حصل فرد فى مـستوى عمرى معين على الدرجة الخام الكلية ١٣٠ فى هذا الاختبار، فـإن معنى ذلك أن درجته تفـوق متوسط مجموعـته العمرية بقـدر انحرافين معياريين.

وتفسر جميع هذه المعايير بنفس الطريقـة التى تفسر بها معايير الدرجات التائية. وعادة تزودنـا أدلة الاختبارات بـجداول تحول فيها الدرجـات الخام مباشـرة إلى هذه المعايير،وتسمى جداول المعايير Tests Norms .

التساعيات المعيارية :

تعد من المعايير التى تعتمد أيضًـا على تحـويل الدرجات المعيارية إلى نظام آخر يختلف فى المتوسط والانحراف المعيارى عن المعايير السابقة، فالتـساعيات المعيارية Stanines هى درجات مـعيارية اعتـدالية متوسطـها ٥ وانحرافهـا المعيارى ٢ تقـريبًا، واشتقت اسمها Stanine من كلمتى Stan اختصار لكلمةStandaradized أى مقنن أو معيارى ، وكلمة nine أى تسعة، وبذلك سميت التساعيات المعيارية.

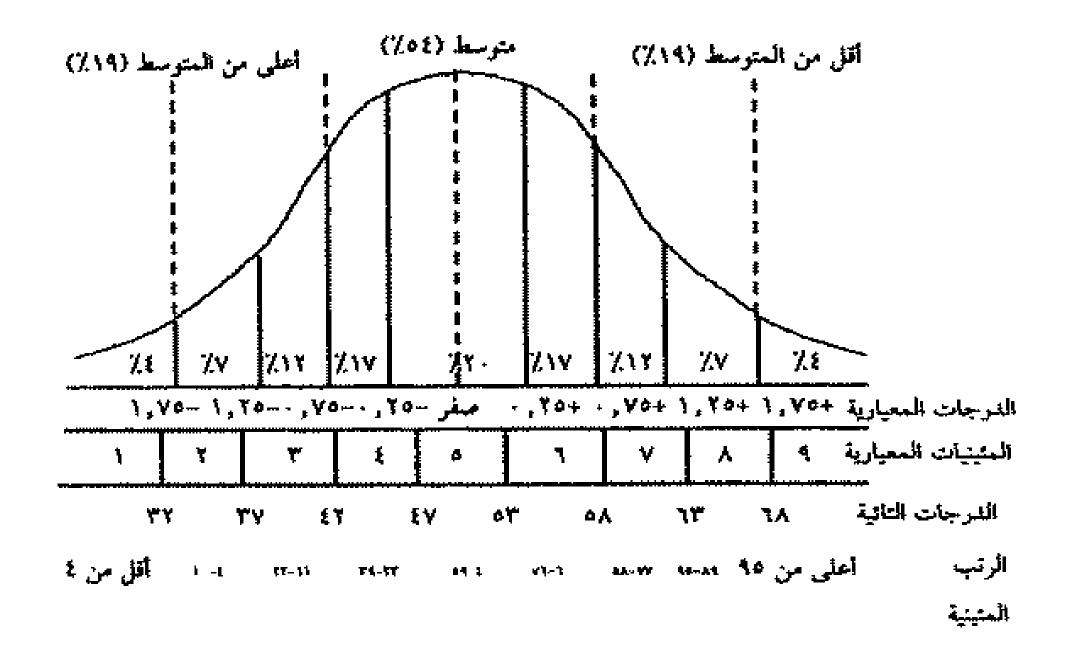
ولم يستخـدم هذا المعيار بكثرة إلا قرب الحرب العالمية الثانيـة، حيث كانت القوات الجوية الأمريكية تـنفذ برنامجًا نفسيًّا يتطلب تطبيق العديد من الاختبارات على آلاف من الجنود، وكان لابد من البـحث عن معيار بسيط يمكن استـخدامه فى تحويل الدرجات الخـام لهذه الاختبارات إلى نـوع آخر من الدرجات يسهـل فهمها وتفـسيرها وتقرير نتائجها.

وقد وجد علماء النفس أن التساعيات المعيارية تفى بذلك، وانتشر استخدام هذا المعـيار بعـد ذلك فى كثيـر من الاختبـارات التربويـة والنفسيـة مرجعـية الجمـاعة أو المعيار،بل أصبح يفضل استخدامه عن غيره من معايير الدرجات المحولة الأخرى. ولا يرجع ذلك فقط إلى سـهولة اشتقاقه واستخـدامه، وإنما إلى مزايا أخرى تتعلق بتفسير الدرجات سوف نوضحها بعد قليل.

ويتضح من اسم هذا المعيار أنه يشتمل على تسع مجموعات من نقاط الدرجات بحـيث تقع نـسبة مـئوية مـعينـة من الأفراد الـمختـبرين فـى كل مجـموعـة من هذه المجـموعات. وبعبارة أخـرى تنحصر نـسبة مئوية مـعينة من الأفراد بين الانـحرافات المعياريـة المختلفة تحت المنحنـى الاعتدالى، وبذلك ينقسم المنحـنى الاعتدالى إلى

تسعة أقسام بواسطة التساعيات المعيارية. ويحد كل قسم منها فترات تمثل نصف انحراف معيارى، ويعيّن لكل قسم الأعداد الصحيحة المرتبة من ١ إلى ٩. أى أن التساعيات المعيارية ليست نقاطًا وإنما فترات، ولذلك فهى تختلف عن الدرجات المحوّلة السابقة.

ويوضح الشكل (٦ ـ ٢) التالى الأقسام التسعة، والنسب المئوية التقريبية تحت المنحنى الاعتدالى، والدرجات المعيارية والتائية والرتب المئينية المناظرة لها:



شكل (٦ ـ ٢) يوضح التساعيات المعيارية والنسب المئوية من الأفراد

والدرجات المعيارية والتائية المناظرة لكل منها

ويتضح من شكل (٦ ـ ٢) أن سعة التساعيات المعيارية من الثانى إلى الثامن متساوية مما يدل على وحدات درجات معيارية متساوية، بينما كل من التساعى الأول والتاسع أكثر اتساعًا من التساعيات الأخرى، ونهايتهما مفتوحة. كما يتضح أن التساعى المعيارى الخامس يمتد من ربع انحراف معيارى يسار المتوسط إلى ربع انحراف معيارى يمين المتوسط، ونقع حوالى ٢٥٪ من الحالات بين هاتين القيمتين. وكذلك نلاحظ أنه كلما اتجهنا إلى أقصى اليمين أو أقصى اليسار تقل النسب المئوية لعدد الحالات، ويمكن اشتقاق هذه النسب المئوية بالطرق الرياضية.

ولتحويل الدرجات الخام إلى تساعيات معيارية ينبغي ترتيب الدرجات ترتيبًا تنازليًّا، ونعيِّن للنسبة ٤٪ العليا من الدرجات التساعي المعياري ٩ ، وللنسبة المئوية ٧٪ التالية في الترتيب التساعي المعياري ٨، وللنسبة المئوية ١٢٪ التالية التساعي المعياري ٧ ، وهكذا.

وبهذا الشكل نكون قد فرضنا التوزيع الاعتدالي على الدرجات الخام أو جعلنا توزيع هذه الدرجات يقترب من التوزيع الاعتدالي، بينما الدرجات المعيارية الخطية أو الدرجات التائية الخطية لا تغيِّر من شكل التوزيع الأصلي للدرجات.

ولكن ينبغي مراعاة أنه إذا كان توزيع درجات سمة معينة لمجموعة من الأفراد ليس اعتداليًا بطبيعته، فإنه لا يجوز إجبار هذا التوزيع ليكون اعتداليًّا. ولكن إذا كان ابتعاد توزيع الدرجات عن الاعتدالية ناجم عن أخطاء القياس، فعندئذ يجوز فرض التوزيع الاعتدالي عليه.

فجميع الأفراد الذين تقع درجاتهم داخل فترة معينة يعيَّن لهم التساعي المعياري المناظر لهذه الفترة. فمثلاً يعيَّن لجميع الأفراد الذين درجاتهم الخام تناظر الدرجات المعيارية المحصورة بين + ٢٥,٠ ، + ٧٥,٠ أو الدرجات التائية المحصورة بين ٥٣، ٥٨ التساعي المعياري ٧ حيث تقع ١٢٪ من الحالات ، وهكذا في بقية الدرجات.

ومن ميزات التساعيات المعيارية أن تساوي فتراتها من التساعي الثاني حتى التساعي الثامن يعني أن التساعي ٨ أفضل من التساعي ٦ بنفس قدر أفضلية التساعي ٥ عن التساعي ٣ ؛ لذلك فإن أداء الأفراد في الاختبارات الفرعية المختلفة الذي يُعبِّر عنه بالتساعيات المعيارية يوضح بدقة الجوانب النسبية لقوتهم وضعفهم في سمات مختلفة. ومما يدل على بساطة هذا المعيار أنه يبين مباشرة المركز النسبي لطالب في ضوء أداء أقرانه في الفرقة الدراسية نفسها.

فعندما يعلم الطالب أن درجته في اختبار ما تقابل التساعي المعياري ٧ أو ٨ أو ٩ فإنه يفهم من ذلك مباشرة أن أداءه فيما يقيسه الاختبار يتميز على أداء الطالب المتوسط في فرقته.

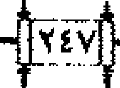
أما إذا كانت درجته تقابل التساعي المعياري ٢ أو ٣ فإن أداءه يكون ضعيفًا نسبيًا. ولذلك يفضل تفسير التساعيات المعيارية تفسيرًا وصفيًا كما هو موضح فيما بين الخطوط المتقطعة في شكل (٦ ـ ٢) فمثلا يمكن تفسير درجات اختبار يقيس المهارات الأساسية في القراءة كالتالي:

التساعي المعياري	وصف الأداء
٩	مرتفع (٤٪)
٨ ، ٧	أعلى من المتوسط (١٩٪)
٦ ، ٥ ، ٤	متوسط (٥٤٪)
٣ ، ٢	أقل من المتوسط (١٩٪)
١	ضعيف (٤٪)

كما يمكن تفسير كل تساعي معياري على حدة من حيث وصفه للأداء، ولكن الوصف في هذه الحالة يكون أكثر تمييزاً، مثل أداء مرتفع جداً، أداء مرتفع، أداء فوق المتوسط، أداء فوق المتوسط بقليل، متوسط، أداء أقل من المتوسط بقليل، أداء أقل كثيراً من المتوسط، أداء ضعيف، أداء ضعيف جداً، ويعتمد كل من التفسيرين على طبيعة الاختبار ودرجة التمييز المرجوة.

وبذلك تفيد التساعيات المعيارية في ترتيب أداء أفراد مجموعة معينة في اختبار واحد أو عدة اختبارات، ومقارنة أدائهم في هذه الاختبارات باستخدام هذا المعيار لمتابعة تقدمهم الدراسي أثناء العام أو في عدة أعوام متتالية. فالتساعي المعياري ٣ مثلا لا يختلف معناه من اختبار إلى آخر، حيث يدل على نسبة مئوية من الأفراد تنحصر درجاتهم بين درجتين معياريتين معينتين.

كما يمكن تعيين أوزان معينة للتساعيات المعيارية للحصول على درجة مركبة يمكن استخدامها في التقويم الشامل لأداء الطالب. وتفيد التساعيات المعيارية أيضًا في الإرشاد والتوجيه التربوي والمهني للطلاب، وبخاصة في نهاية بعض المراحل التعليمية حيث تتخذ قرارات تتعلق بالتخصصات أو المناهج الدراسية المناسبة للطلاب وكذلك الاختبارات المهنية. إذ يمكن استخدامها في التمثيل البياني لمستوى أداء الطالب في مجموعة من الاختبارات المتنوعة المناسبة لهذه القرارات، بحيث تتضح جوانب القوة والضعف للطالب مما يساعده في اتخاذ قراره استناداً إلى بيانات موضوعية بدلاً من اعتماده على أحكامه الذاتية.



وعلى الرغم من مزايا التساعيات المعيارية سواء فى إعدادها أو تقرير نتائجها باستخدام رقم صحيح واحد، إلا أن هناك بعض أوجه قصور تنجم عن تفسير درجات الاختبارات فى ضوء هذا المعيار.

ويتلخص ذلك فى أن جميع الأرقام المناظرة للتساعيات المعيارية تمثل مدى متسع نسبيًا من مقدار السمة المقاسة كما رأينا. ولذلك يصعب تحديد مقدار النمو الذى حدث فى سمة من السمات كالتحصيل أو الذكاء مثلا باستخدام هذا المعيار، فالدرجة الخام التى يحصل عليها طالب فى اختبار تحصيلى معين يمكن أن تزداد بقدر ١٥ نقطة إذا طبق عليه الاختبار قبل عملية التعليم وعقب الانتهاء منها، ومع ذلك تظل درجته فى كل من مرتى التطبيق مقابلة لنفس التساعى المعيارى نظرًا لأن مدى التساعيات متسع كما ذكرنا.

وإذا افترضنا أن الطالب (أ) حصل على الدرجة الخام ٨٤ فى التطبيق الأول لأحد الاختبارات ، وأن هذه الدرجة تعد أعلى درجة مقابلة للتساعى المعيارى ٦ ، وأن الطالب (ب) حصل على الدرجة ٨٥ فى التطبيق الأول للاختبار نفسه، وأن درجته تعد أقل درجة مقابلة للتساعى المعيارى ٧، فإذا أعيد تطبيق الاختبار على كل من الطالبين وحصلا على الدرجتين الخام ٨٥ ، ٨٤ على الترتيب واستخدمنا جدول المعايير نفسه لتفسير هذه الدرجات ، فإننا نلاحظ أن الطالب (أ) قد ارتفعت درجته من ٨٤ إلى ٨٥ أى من التساعى المعيارى ٦ إلى التساعى المعيارى ٧ ، بينما انخفضت درجة الطالب (ب) من ٨٥ إلى ٨٤ أى من التساعى المعيارى ٧ إلى التساعى المعيارى ٦ .

وبذلك يبدو أن الطالب (أ) قد حدث له نمو على عكس الطالب (ب) ، مع أن هذا التغير فى التساعيات المقابلة للدرجات الخام حدث نتيجة تغير الدرجات الخام نقطة واحدة ارتفاعًا أو انخفاضًا.

كذلك لا نستطيع باستخدامها التمييز بين الأفراد الذين تفوق درجاتهم فى أحد الاختبارات التساعى المعيارى ٩ أو تقل درجاتهم عن التساعى المعيارى ١ ، نظرًا لأن نهاية كل منهما مفتوحة.

لذلك اقترح علماء القياس جعل عدد التساعيات ١١ بدلا من ٩ لمزيد من دقة التمييز، وأطلقوا على المعيار الجديد اسم المعيار ذو الأحد عشر قسمًا **Sta - eleven**، ويُقسم هذا المعيار كل من التساعى ١ إلى قسمين صفر ، ١ ، والتساعى ٩ إلى قسمين ٩ ، ١٠ بحيث تقع نسبة ١ ٪ من الدرجات فى كل من القسمين صفر ، ١٠ تحت المنحنى الاعتدالى.

ونلاحظ مما سبق أن معايير الاختبارات والمقاييس التى تعتمد على الدرجات المحوّلة تختلف فيما بينها فى القيم الاختيارية لكل من متوسط درجات الجماعة المرجعية والانحراف المعيارى. ويمكن تلخيصها فى الجدول (٦ ـ ١) التالى:

الدرجات المحولة	خصائص الجماعة المرجعية
الدرجة المعيارية	س = صفر ، ع = ١
الدرجة التائية	س = ٥٠ ، ع = ١٠
معيار اختبار الاستعداد الأكاديمى	س = ٥٠٠ ، ع = ١٠٠
معيار اختبار التصنيف العام	س = ١٠٠ ، ع = ٢٠
نسبة الذكاء الانحرافية	س = ١٠٠ ، ع = ١٥أو ١٦
التساعيات المعيارية	س = ٥ ، ع = ٢

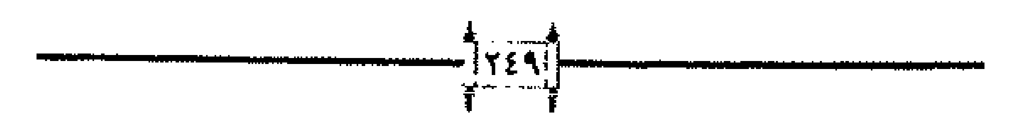
جدول (٦ ـ ١) ملخص معايير الاختبارات التى تعتمد على

الدرجات المحولة

ومما هو جدير بالذكر أن الدرجات المحوَّلة المستخدمة في الاختبارات والمقاييس التربوية والنفسية لا تقتصر فقط على ما هو موضح بجدول (٦ ـ ١) ، وإنما يمكن التوصل إلى درجات محولة أخرى باختيار قيم مناسبة لكل من المتوسط والانحراف المعيارى للدرجات الخام للجماعة المرجعية، وإجراء التحويل باستخدام الصيغة (٣ ـ ٧) المذكورة فى الفصل الثالث.

ثانيا : معايير المئينيات والرتب المئينية :

تعد المئينيات والرتب المئينية نوعًا آخر من معايير الاختبارات والمقاييس التى تستخدم فى أغراض انتقاء وتسكين الأفراد، ولا تعتمد على التحويلات الخطية التى وضحناها فيما سبق. فالمئينيات Percentiles هى نقط على توزيع الدرجات تقع دونها نسبة مئوية معينة من هذه الدرجات، أما الرتب المئينية Percentile Ranks فإنها تعين الموقع النسبى للفرد أو النسبة المئوية من درجات أقرانه التى تقل عن درجته. فالمئينى ٥٥ مثلاً يعنى النقطة التى تقع دونها ٥٥٪ من الدرجات فى التوزيع، وهذا لا يعنى أن الفرد الذى درجته الخام فى الاختبار تقابل المئينى ٥٥ أنه حصل على ٥٥ درجة، إذ ربما يكون قد حصل على ٧٠ درجة ، ولكنها تعنى أن ٥٥٪ من الافراد



المختبرين حصلوا على درجات تقل عن ٧٠ درجة. لذلك فإن الدرجة الخام ٧٠ تقابل المئيني الذى رتبته ٥٥ ، أى أن ٥٥ هي الرتبة المئينية.

وقد سبق أن أوضحنا بالعديد من الأمثلة في الفصل الثالث كيفية حساب المئينيات والرتب المئينية ومزاياها وعيوبها.

ونظرا لسهولة حساب وفهم هذا النوع من المعايير فإنه شائع الاستخدام في كثير من الاختبارات التربوية والنفسية المنشورة. وتزودنا أدلة هذه الاختبارات بجداول معايير مئينية يمكن الرجوع إليها لتفسير الدرجات الخام التى يحصل عليها الأفراد فى الاختبار. وأحيانًا تحدد معايير مئينية مستقلة للمهن، والفرق الدراسية، والأعمار المختلفة، وكذلك معايير منفصلة لكل من الذكور والإناث.

ويمكن استخدام هذه الجداول في إجراء مقارنات داخل الفرد وبين الأفراد، فمثلاً هناك جدول معايير المئينيات لاختبار الاستعدادات الفارقة Differential Aptitude Test (DAT) الذى يشتمل على ثمانية اختبارات تقيس الاستدلال اللفظى، والقدرة العددية، والاستدلال المجرد، والاستعداد الكتابى، والاستدلال الميكانيكى، والعلاقات المكانية، والهجاء، والنحو.

ويوضح جدول (٦ ـ ٢) التالى الرتب المئينية المناظرة لدرجات أحد الطلاب فى كل من هذه الاختبارات الفرعية:

	الدرجة الخام	الرتبة المئينية
الاستدلال اللفظى	٣٣	٦٥
القدرة العددية	٢٧	٧٠
الاستدلال المجرد	٣٤	٤٥
الاستعداد الكتابى	٤١	٤٠
الاستدلال الميكانيكى	٤٠	٢
العلاقات المكانية	٣٥	٥٥
الهجاء	٦٢	٥٥
النحو	٢٩	٦٥

جدول (٦ ـ ٢) يوضح الرتب المئينية المناظرة لدرجات أحد الطلاب في اختبارات الاستعدادات الفارقة

وبالنظر إلى جدول (٦ ـ ٢) يمكن مقارنة الرتب المئينية للطالب في الاختبارات الفرعية، كما يمكن مقارنة درجاته في كل اختبار فرعي بدرجات الجماعة المرجعية. ونلاحظ أن رتبته المئينية المناظرة لدرجته الخام في القدرة العددية ٧٠، أي أنه يتفوق على ٧٠ ٪ من أقرانه في هذه القدرة، كما أنها أعلى الرتب المئينية المناظرة لدرجاته في الاختبارات الفرعية المختلفة.

وهنا ربما يتساءل مستخدم الاختبار عن كيفية الحكم على رتبة مئينية معينة ولتكن ٨٥ مثلاً، أي هل هذه رتبة مئينية مرتفعة؟ والإجابة عن ذلك تعتمد على الغرض من استخدام الاختبار وطبيعة الجماعة المرجعية، فإذا كان الغرض انتقاء أفضل الأفراد للقيام بمهام معينة أو للالتحاق بدراسات تتطلب كفايات متميزة، فإن الرتبة المئينية ٨٥ ربما تبدو غير مقبولة.

أما إذا كان الغرض تحديد مكانة الفرد بين أقرانه في صف دراسي معين، وكانت خلفية الجماعة المرجعية وخبراتها مشابهة لخبرات الفرد المختبر، فإن هذه الرتبة المئينية ربما تبدو جيدة. فالرتبة المئينية المقابلة لدرجة خام معينة لا تشير إلى ما يمتلكه الفرد من معارف أو مهارات أو خبرات تتعلق بمحتوى الاختبار، وإنما تحدد فقط مكانته النسبية بين أقرانه. فالفرد الذي رتبته المئينية ٥٠ في اختبار يقيس المهارات الحسابية لا ينبغي تفسيرها على أنه ضعيف التحصيل، وإنما تعني أن درجته في الاختبار تقل عن ٥٠٪ من درجات أقرانه الذين طبق عليهم الاختبار نفسه.

وكما في معظم معايير الاختبارات، لا يجب عند تفسير الرتب المئينية اعتبارها قيمًا مطلقة أو غاية في الدقة؛ نظرًا لأن الأفراد الذين يحصلون على درجات تقابل رتبتين مئينيتين معينتين مثل ٤٢ ، ٤٤ تكون درجاتهم متقاربة.

لذلك يفضل ناشرو الاختبارات تدوين المئينيات في أدلة الاختبارات بحيث تتضمن مدى معين Percentile Bands بدلا من اعتبارها نقطًا محددة وذلك مراعاة لأخطاء القياس وجعل تفسيرها أكثر تحفظًا. فبدلاً من القول بأن الدرجة الخام ٦٢ في اختبار ما تناظر المئيني ٥٧ يمكن اعتبارها مناظرة للمدى المئيني ٥٦ ـ ٥٨ مثلاً.

ولعل من ميزات المئينيات والرتب المئينية أنه يمكن الاستناد إليها في تفسير جميع الدرجات المحولة التي أوضحناها فيما سبق والمبينة في شكل (٦ ـ ٢). فالدرجة التائية ٣٧ مثلاً تقابل الرتبة المئينية ١١ ، والتساعي المعياري ٤ يقابل الرتب المئينية المحصورة بين ٢٣ ، ٣٩ وهكذا.

ثالثا : معايير الارتقاء :

يطَّرد نمـو الفرد في كثير من السمات الإنسانية بمرور الزمن بطريقة منظمة، فتحصيل الطالب يرتقي بانتقاله من فرقة دراسية إلى أخرى، وقدراته العقلية تنمو، ومهارات الأفراد تتحسن من عام إلى آخر، وهكذا. فالنزعة الارتقائية تميز مختلف جوانب سلوك الإنسان، لذلك اهتم علماء القياس ببناء معايير ارتقائية يمكن تفسير درجات الاختبارات استناداً إليها بحيث يمكن مقارنة أداء الفرد بمتوسط أداء المستويات النمائية المختلفة. ولعل أكثر هذه المعايير استخداماً وبخاصة في الاختبارات التحصيلية وبعض اختبارات القدرات معايير العمر الزمني Age Norms، ومعايير الفرق الدراسية Grade Norms ؛ وذلك لأنه يسهل على مستخدمي الاختبارات فهمها وتفسيرها.

معايير العمر الزمني:

تسمى هذه المعايير أحياناً العمر التعليمي Educational Age، أو مكافئ العمر الزمني Age Equivalent. وتفيد هذه المعايير إذا توقعنا أن نمو الفرد يحدث نتيجة لعوامل النضج.

فإذا كان محمد تلميذاً في الصف الرابع الابتدائي وطوله ١٣٨ سم فربما نعتبره أكثر طولاً بالنسبة لتلاميذ هذا الصف، غير أنه ربما يكون من الأفضل مقارنته بالأطفال الذين في نفس مستواه العمري، بدلا من مقارنته بتلاميذ فرقته. فإذا كان عمره الزمني ٩سنوات ، ١١ شهراً ، ووسيط الطول لأطفال هذا العمر (الجماعة المرجعية) ١٤٠سم فإنه يعد أقصر من وسيط طول أقرانه. لذلك إذا كان لدينا معايير العمر الزمني أو معايير مكافئ العمر للأطوال، فإنه يمكننا تحديد المجموعة العمرية المناظرة لطول محمد، وبالتالي يمكن تقييم طوله.

فلكي نحدد مكافئ العمر لطول محمد، ينبغي أن نُحدد جماعة مرجعية وسيط طولها ١٣٨ سم . فإذا افترضنا أن عمر أفراد هذه المجموعة ٩ سنوات، ٥ شهور، فإن مكافئ عمر محمد حينئذ يكون ٩ سنوات، ٥ شهور ، وتكتب (٩ ـ ٥).

وقد استخدمنا الشرطة الأفقية للفصل بين الأعوام والشهور بدلاً من وضع فاصلة، وذلك لأن معايير مكافئ العمر تعتمد على نظام ١٢ شهراً، أما معايير العمر الزمني التي تعتمد على نظام ١٠ شهور فتستخدم فاصلة بين الأعوام والشهور ، أي (٩ ، ٥) مثلاً.

وللحصول على معيار مكافئ العمر فى خاصة معينة من خصائص الفرد، نطبق الاختبار المتعلق بهذه الخاصة على مجموعات من الأفراد ممثلة لمستويات عمرية مختلفة، ونوجد قيمة وسيط توزيع الدرجات فى كل مستوى عمرى على حدة. ويمكن عندئذ تفسير الدرجات الخام على أنها تمثل مستوى أداء مناظر لوسيط أداء عمر زمنى معين.

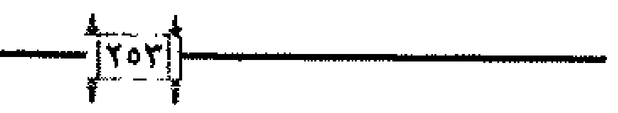
وتُفسر الدرجات الخام فى بعض اختبارات القدرات العقلية على أنها تُمثل العمر العقلى Mental Age ، وفى اختبارات القراءة مثلا تمثل العمر القرائى Reading Age وهكذا.

فإذا كان متوسط درجات جميع التلاميذ الذين يبلغون من العمر عشرة أعوام فى أحد الاختبارات ٤٢، فإن أى تلميذ يحصل على الدرجة ٤٢ فى هذا الاختبار يكون مكافئه العمرى ١٠ أعوام ، بغض النظر عن عمره الفعلى. ويفترض بعض مؤلفى الاختبارات أن درجات الاختبار تزيد زيادة منتظمة من مستوى عمرى إلى آخر؛ لذلك يُعدون معايير للمستويات العمرية التى تنحصر بين الأعوام المختلفة ، أى ١٠ أعوام ، ٤ أشهر مثلاً. وهنا يمكن إضافة ثلث الفرق بين متوسط درجات التلاميذ الذين تبلغ أعمارهم ١٠ أعوام ، ١١ عامًا إلى الدرجة التى حصل عليها التلميذ فى الاختبار، ويكون مكافئها العمرى ١٠ أعوام ، ٤ شهور.

ومعايير العمر الزمنى يمكن تفسيرها بسهولة إذا ارتبط ما يقيسه الاختبار بالعمر الزمنى مثل القدرة العقلية، والقدرة على القراءة ، وكثير من المهارات الأساسية التى تهدف المدرسة الابتدائية لتنميتها لدى التلاميذ.

غير أن معايير مكافئ العمر يشوبها بعض العيوب التى ينبغى مراعاتها عند تفسير درجات الاختبارات التى تستند إليها.

ومن بين هذه العيوب صعوبة تفسير الدرجات المرتفعة أو المنخفضة بدرجة كبيرة ؛ نظرًا لأن هذه المعايير تقتصر على المستويات المتوسطة للدرجات الخام. فإذا ما أردنا تفسير الدرجات التى تخرج عن هذا النطاق ينبغى إجراء عملية استكمال Exter-polation للخط النمائى البيانى اعتماداً على فرضيات ربما يصعب تبريرها. فإذا كانت المهارات التى تقيسها الاختبارات تنمو نموًّا منتظمًا بزيادة العمر، فإن عملية الاستكمال البيانى ربما تكون ممكنة، أما إذا تباين معدل النمو من عام إلى آخر، فعندئذ يصعب إجراء ذلك. فمعدل نمو فهم المقروء مثلاً لدى الطلاب الذين يبلغون من العمر ١٧ أو ١٨ عامًا أقل منه لدى التلاميذ الذين يبلغون من العمر ٧ أو ٨ أعوام.



لذلك فإن تفسير فرق درجات عمر عام واحد يختلف باختلاف المستويات العمرية . ويفضل استخدام معايير مكافئ العمر في المرحلة الابتدائية ، حيث إن فرضية تساوي وحدات النمو يمكن أن تتحقق في هذه المرحلة بدرجة أفضل من غيرها من المراحل.

كما أن هذا النوع من المعايير يعتمد على السياسات التعليمية المتعلقة بالحد الأدنى لأعمار التلاميذ الذين يلتحقون بالمدارس ، وأساليب انتقالهم إلى صفوف أعلى أو تسريعهم.

فإذا كانت الجماعة المرجعية المستخدمة لبناء معايير عمر ١٠ أعوام اشتملت على تلاميذ انتقلوا بالأساليب المعتادة إلى الصفوف الأعلى، فإن هذه المعايير سوف تختلف عن المعايير التي تعتمد على تلاميذ انتقلوا بأساليب التسريع؛ لذلك ينبغي تحديد نسبة كل مجموعة من هؤلاء التلاميذ في مستوى عمري معين تحديداً دقيقًا لكي تفسر معايير مكافئ العمر الزمني تفسيراً معقولاً.

معايير الفرق الدراسية :

يطلق على هذا النوع من المعايير «معايير مكافئ الفرق الدراسية Grade Equivalent» ويشبه إلى حد كبير معايير مكافئ العمر. وهذه المعايير تناسب الاختبارات التحصيلية التي يستخدمها المعلمون في الصفوف الدراسية المختلفة وبخاصة في مرحلة التعليم الابتدائي. فالدرجة التي يحصل عليها الطالب في الاختبار يتم مقارنتها بوسيط درجات الجماعة المرجعية في فرقة دراسية معينة. ونحصل على هذه المعايير بنفس طريقة معايير مكافئ العمر الزمني، غير أننا هنا نطبق الاختبار على جماعة مرجعية ممثلة لمختلف الفرق الدراسية وإيجاد وسيط درجات كل فرقة على حدة. ونستخدم قيمة وسيط كل من هذه الفروق في تفسير الدرجات الخام، أي نحدد الفرقة الدراسية المكافئة لدرجة خام معينة في الاختبار.

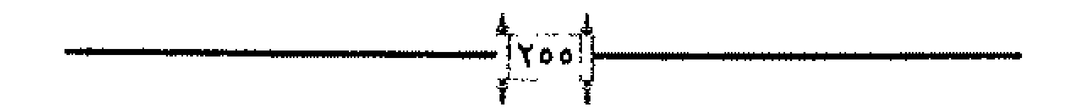
ويوضح جدول (٦ ـ ٣) معايير مكافئ العمر الزمني ومعايير الفرق الدراسية لاختبار افتراضي في فهم معاني الكلمات لتلاميذ الصفوف الأولى للمرحلة الابتدائية.

معانى الكلمات		
الدرجة	مكافئ العمر الزمنى	مكافئ الفرقة الدراسية
١	٧ - ١	١,٩
٢	٧ - ٢	٢,٠
٣	٧ - ٣	٢,١
٤	٧ - ٣	٢,١
٥	٧ - ٥	٢,٢
٦	٧ - ٥	٢,٢
٧	٧ - ٦	٢,٣
٨	٧ - ٧	٢,٤
٩	٧ - ٨	٢,٥
١٠	٧ - ١٠	٢,٦
١١	٧ - ١١	٢,٧
١٢	٨ ـ صفر	٢,٨
١٣	٨ ـ صفر	٢,٨
١٤	٨ ـ ٢	٢,٩
١٥	٨ ـ ٣	٣,٠

جدول (٦ ـ ٣) يوضح معايير مكافئ العمر الزمنى، ومكافئ

الفرق الدراسية لاختبار افتراضي في فهم معانى الكلمات

ويتضح من جدول (٦-٣) أن الدرجة الخام (١) تناظر مكافئ العمر الزمنى (٧-١) أى سبعة أعوام وشهر واحد، كما تناظر مكافئ الفرقة الدراسية (١,٩) أى الصف الأول والشهر التاسع من العام الدراسي، وهكذا فى بقية الدرجات.



وتعتمد هذه المعايير اعتماداً أساسيًا على أساليب انتقاء المدارس والتلاميذ والسياسات التعليمية التى سبق أن أشرنا إليها، ويتم تدوين هذه المعايير على أساس عشرة أشهر وليس ١٢ شهراً، كما فى معايير مكافئ العمر، أى تتراوح مثلاً بين (٥،صفر) ، (٥ ـ ٩) لفترات شهرية للعام الدراسى . ويفضل وضع فاصلة بين الفترات، فمثلاً الدرجة ٦,٧ تعنى الصف السادس ، والشهر السابع من العام الدراسى . والشكل (٦ ـ ٣) التالى يوضح الأجزاء الكسرية لمكافئ الفرق الدراسية :

شهور العام الدراسى

سبتمبر	أكتوبر	نوفمبر	ديسمبر	يناير	فبراير	مارس	أبريل	مايو	يونية
,٠	٠,١	٠,٢	٠,٣	٠,٤	٠,٥	٠,٦	٠,٧	٠,٨	٠,٩

الجزء الكسرى لمكافئ الفرق الدراسية

شكل (٦ ـ ٣) يوضح الأجزاء الكسرية لمكافئ الفرق الدراسية

وكثيراً ما يخطئ البعض فى تفسير مكافئ الفرق الدراسية، حيث يظنون أن هذا المعيار يشير إلى الفرقة الدراسية التى ينبغى أن يلتحق بها التلميذ. فإذا كان مكافئ الفرقة الدراسية لتلميذ فى الصف الخامس الابتدائى ٧,٣ فهذا لا يعنى أنه يجب نقل هذا التلميذ إلى الفرقة السابعة، أو أنه ينبغى إعطاؤه واجبات مدرسية مناسبة للفرقة السابعة، ولكنها تعنى فقط أن أداء هذا التلميذ يفوق متوسط أداء تلاميذ الصف الخامس.

ومن الأخطاء الأخرى الشائعة فى تفسير درجات هذا المعيار مقارنة مكانة التلميذ فى أحد الاختبارات بمكانته فى اختبار آخر، فالانحراف المعيارى يختلف اختلافًا ملحوظًا من اختبار إلى آخر، وهذا يجعل المقارنة غير ممكنة كما أوضحنا فى الفصل الثالث.

وتتميز معايير مكافئ الفرق بنفس ميزات معايير مكافئ العمر ويشوبها نفس عيوبها، فاستخدام معايير مكافئ الفرق الدراسية يكون مناسبًا فقط إذا كانت الخاصة التى يقيسها الاختبار تتباين تباينًا منتظمًا من فرقة دراسية إلى أخرى، وافتراض تكافؤ الخبرات المدرسية، لذلك يفضل استخدامها فى المرحلة الابتدائية.

والخلاصة أن معايير الارتقاء تفيد فى تفسير درجات الاختبارات بوحدات يمكن فهمها ببساطة سواء كانت مكافئ العمر أو مكافئ الفرق الدراسية، ويمكن باستخدامها مقارنة أداء الفرد مباشرة بأداء أقرانه، ومقارنة أداؤه فى اختبارات مختلفة لتعرف مقدار تقدمه أو نموه بمرور الزمن. ولكنها لا تناسب الخصائص التى لا تتغير بانتظام بمرور الزمن أو من فرقة دراسية إلى أخرى، وتتأثر بتكوين الجماعة المرجعية وخلفياتها وخبراتها. كما أنه ينبغى إجراء عملية استكمال للخط البيانى من عمر إلى آخر أو من فرقة دراسية إلى أخرى فيما يتعلق بالشهور، مما يتطلب افتراضات معينة ويقلل من دقة الدرجات وبخاصة أن وحدات هذه المعايير تختلف من عمر إلى آخر أو من فرقة دراسية إلى أخرى.

لذلك يفضل استخدام الرتب المئينية داخل كل فرقة دراسية بدلاً من وسيط درجات الفرق، فنقول مثلاً أن التلميذ (أ) ربما تكون الرتبة المئينية المقابلة لدرجته فى الاختبار ٩٢ بالنسبة لتلاميذ الصف الخامس الابتدائى.

رابعًا: الصفحة النفسية أو التربوية (البروفيل النفسى أو التربوى):

تشتمل بعض الاختبارات على عدة اختبارات فرعية تقيس سمات أو قدرات أو مهارات متعددة تشكل جميعها بطاريات من الاختبارات Tests Batteries وتستخدم فى أغراض محددة مثل الانتقاء أو التسكين أو التشخيص. ومن أمثلة ذلك بطارية اختبارات الاستعداد الأكاديمى (SAT) ، واختبار ويكسلر لذكاء الراشدين (WAIS)، واختبار ويكسلر لذكاء الأطفال(WISC) ، واختبار الشخصية متعدد الأوجه (MMPI) ، واختبار الميول المهنية (SVIB) ، واختبار الاستعدادات الفارقة (DAT)، وبطارية اختبارات الاستعداد العام (GATB) ، وبطارية اختبارات المهارات الأساسية الشامل(CTBS) وبطارية اختبارات السجل الأكاديمى (GRE) وغيرها من بطاريات الاختبارات التى سنوضحها فى الفصول التالية، فإذا طبقت أى من هذه البطاريات على أحد الأفراد، فإنه يحصل على عدة درجات وفقًا لعدد الاختبارات التى تشتمل عليها البطارية.

ويكون الغرض من استخدام هذه البطاريات ليس فقط معرفة أدائه النسبى فى كل اختبار على حدة، وإنما أيضًا مقارنة أدائه فى الاختبارات المختلفة.

ولتوضيح الصورة العامة للأداء وتيسير عملية المقارنة يزود ناشرو الاختبارات مستخدمى هذه البطاريات ببطاقة أو ورقة رسم بيانى معدة بطريقة خاصة تفيد فى التمثيل البيانى لدرجات الاختبارات التى تشتمل عليها بطارية معينة؛ وذلك لتوضيح موقع الفرد

أو مستوى أدائه فى هذه الاختبارات التى تمثل أبعادًا نفسية أو معارف ومهارات تربوية معينة ، وتحديد جوانب القوة والضعف. ويسمى هذا التمثيل البيانى النفسى أو التربوى الصفحة النفسية (البروفيل النفسى) Psychological Profile ،أو الرسم البيانى التربوى (البروفيل التربوى) Educational Profile .

ويمكن إعداد الصفحة النفسية لمجموعة من درجات الاختبارات لفرد معين، حيث تمثل الاختبارات على المحور الأفقى، ويمثل مقدار كل درجة بنقطة على ارتفاع مناسب مواز للمحور الراسى. ولكى نلاحظ النقط بسهولة يمكن توصيل هذه النقط بخطوط مستقيمة مما يعطى شكل البروفيل، غير أن هذه الخطوط لا تعنى شيئًا حيث إن ما يهمنا هو النقط المناظرة لدرجات الاختبارات؛ لذلك ربما يفضل عدم توصيل النقط فيما بينها.

ويوضح شكل (٦ ـ ٤) التالى صفحة نفسية لدرجات أحد الطلاب فى بطارية اختبارات الاستعدادات الفارقة مستمدة من دليل هذه البطارية الذى أصدرته المؤسسة السيكولوجية الأمريكية عام ١٩٦٣.

شكل (٦ ـ ٤) يوضح صفحة نفسية لدرجات طالب فى اختبار الاستعدادات الفارقة

ويتضح من شكل (٦ ـ ٣) أن الصفحة النفسية تحدد المئينيات والدرجات المعيارية المقابلة للدرجات الخام للطالب فى الاختبارات الفرعية الثمانية التى تشتمل عليها بطارية اختبارات الاستعدادات الفارقة، وتمثل الأعمدة الرأسية هذه الدرجات.

والمئينيات والدرجات المعيارية المقابلة لدرجات جميع اختبارات البطارية مستمدة من جماعة مرجعية واحدة محددة الخصائص.

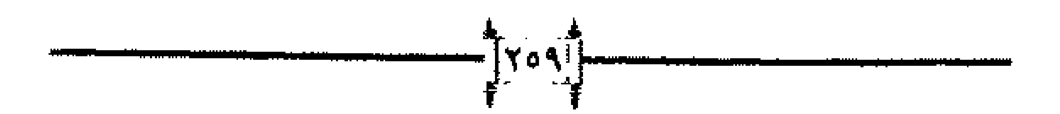
وهناك ثلاثة خطوط أفقية تمتد من النقط المقابلة لكل من المئينى ٢٥ ، والمئينى ٥٠، والمئينى ٧٥ ، أى الإرباعيات الأول والثانى والثالث. وذلك ليتسنى مقارنة درجات الطالب فيما بينها بمجرد النظر ، وكذلك مقارنتها بدرجات الطلاب الآخرين. وتفيد هذه الصفحة النفسية فى تحديد جوانب القوة والضعف فى استعدادات الطالب، والاستناد إليها فى التوجيه والإرشاد التربوى.

فإذا نظرنا إلى الصفحة النفسية للطالب الموضحة بشكل (٦ ـ ٤) نلاحظ أنه حصل على أعلى الدرجات فى الاستعداد الميكانيكى حيث تقابل الرتبة المئينية ٩٥، تليها درجته فى الاستعداد المكانى المقابلة للرتبة المئينية ٨٧ تقريبًا، وحصل على أقل الدرجات فى الاستعداد الكتابى، حيث تقابلها الرتبة المئينية ٢٥ أى الإرباعى الأول، وبذلك تتبين جوانب قوته وضعفه.

وتتميز الصفحات النفسية أو الرسوم البيانية التربوية بخصائص ثلاث هى: المستوى Level ، والانتشار Scatter ، والشكل Shape . «فالمستوى» يعنى الارتفاع العام للأعمدة فى الرسم البيانى وهو يدل على مستوى الأداء العام للفرد أو المتوسط العام لدرجاته المعيارية فى الاختبارات الفرعية، و«الانتشار» يعنى درجة تشتت مختلف النقط التى تمثل الدرجات حول الموقع المتوسط (المستوى) لدرجات جميع الاختبارات الفرعية، ويدل ذلك على مدى اختلاف درجات الفرد فيما بينها. أما «الشكل» فيعنى نمط الاستعدادات المتمثل فى ترتيب درجات الاختبارات التى حصل عليها الفرد.

وهذه الخصائص يمكن أن تختلف إلى حد ما فيما بينها كل على حدة، إذ يمكن أن تتفق صفحتان نفسيتان لفردين فى «المستوى» و«الانتشار»، ومع ذلك تختلف كل منهما عن الأخرى فى «الشكل» ، أى فى ارتفاع وانخفاض النقط التى تمثل درجاتهم فى اختبارات البطارية.

ولا تقتصر الصفحات النفسية أو الرسم البيانى التربوى Profiles على نوع واحد، وإنما تتعدد أنواعها وطرق تمثيلها. فالنوع الذى قدمناه يمثل درجات بطارية اختبارات تطبق على فرد واحد فى وقت واحد لإعطاء صورة إستاتيكية مستعرضة



لدرجاته. بيد أنه يوجد نوع آخر يسمى الصفحات النفسية الارتقائية أو الطولية Longitudinal Profiles حيث تمثل درجات الاختبار نفسه أو عدة اختبارات في أزمنة مختلفة لتوضيح نمو الفرد في سمة أو سمات متعددة. فمثلاً يمكن تمثيل درجات تلميذ في اختبارات القراءة والحساب والمعلومات العامة من الصف الثالث إلى الصف السادس في رسم بياني تربوي Educational Profile للإفادة منها في تعرف التقدم الدراسي للتلميذ وتوجيهه توجيهًا تربويًا سليمًا.

وعادة يوضح مؤلف بطارية الاختبارات أمثلة لصفحات نفسية لبعض الحالات وكيفية استخدامها في الإرشاد والتوجيه التربوي أو النفسي لهذه الحالات للاسترشاد بها في استخلاص المعلومات من صفحة نفسية أخرى، ويستخدم الحاسوب في طباعة وتنسيق هذه الصفحات.

غير أنه يجب مراعاة أن الفرق بين درجتين في اختبارين مختلفين يتأثر بأخطاء القياس المتعلقة بكل منهما، ولذلك فإن ثبات هذا الفرق يكون أقل من ثبات درجات كل اختبار على حدة، كما سبق أن أوضحنا في الفصل الرابع. ونظراً لأن الصفحات النفسية أو الرسوم البيانية التربوية تعكس الفروق بين درجات اختبارات فرعية مختلفة، فإنه ينبغي الحيطة قبل أن نعزى هذه الفروق إلى الاختلافات بين درجات الاختبارات الفرعية التي تمثلها الصفحة النفسية أو الرسم البياني التربوي وبخاصة إذا كان عدد مفردات هذه الاختبارات قليل مما يؤدى إلى انخفاض ثبات درجاتها.

كما يجب مراعاة بعض العوامل الأخرى لكي لا تقدم الصفحة النفسية صورة غير صادقة عن سمات الفرد. إذ ينبغي أن تُستخدم الجماعة المرجعية نفسها في إعداد معايير الاختبارات الفرعية، أما إذا استخدمت جماعات مرجعية مختلفة، فإننا لا نستطيع المقارنة المباشرة بين هذه الدرجات .

كما ينبغي أن تمثَّل درجات مجموعة الاختبارات في الرسم البياني بميزان أو تدريج واحد، وأن يعبر عن الدرجات بالرتب المئينية أو أي نوع من أنواع الدرجات المحولة المقابلة لها. كما يفضل استخدام معيار لا يضخم الفروق الضئيلة أو يصغر الفروق الكبيرة. ويمكن التغلب على ذلك إلى حد ما بأن تتضمن الصفحات النفسية مؤشراً يبين مقدار أخطاء القياس لكي لا يساء تفسير الدرجات.

وقد سبق أن أوضحنا أنه يمكن التغلب على ذلك أيضًا بأن تمثل الدرجات بفترات، أي تنحصر في مدى معين Bands بدلاً من تمثيلها بنقط أي قيم مضبوطة، وتكون الفروق بين الدرجات دالة إذا لم يحدث تداخل بين الفترات.

ومن هنا يتضح أن المعايير Norms تُعد أساسًا لتفسير درجات الأفراد فى الاختبارات والمقاييس مرجعية الجماعة أو المعيار، فتحويل الدرجة الخام التى يحصل عليها الفرد إلى نوع من أنواع المعايير التى أوضحناها تمكننا من تفسير هذه الدرجة فى ضوء المركز النسبى للفرد بين أقرانه المماثلين له فيما يقيسه الاختبار.

وكذلك تمكّننا المعايير من المقارنة بين أداء الفرد فى مجالات متعددة يتم قياسها بواسطة بطارية من الاختبارات حيث يمكن تمثيل درجاتها برسم بيانى تربوى أو بصفحة نفسية تداريجها موحدة، وتفيد الأنواع المختلفة من المعايير التى أوضحناها فى أغراض متعددة، ويمكن أن يتضمن الاختبار أكثر من نوع منها. فقد لاحظنا أن المئينيات يمكن أن تستخدم مع الدرجات المحولة،أو مع مكافئ العمر، أو مكافئ الفرقة الدراسية فى اختبار واحد.

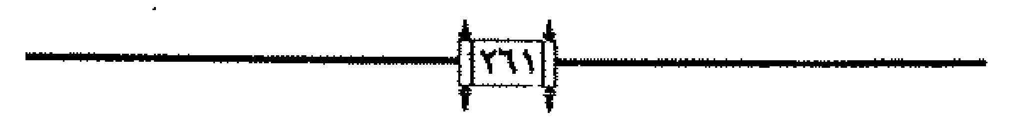
غير أنه فى جميع الحالات ينبغى مراعاة خصائص الجماعة المرجعية التى تُستمد منها المعايير، ومدى تشابهها مع خصائص الأفراد الذين ستُتخذ قرارات بشأنهم فى ضوء هذه المعايير، وأساليب انتقاء أفراد هذه الجماعة، وتاريخ تطبيق الاختبار عليهم، ومقدار أخطاء القياس المتعلقة بأساليب المعاينات، وغير ذلك من الاحتياطات الواجب مراعاتها عند تفسير معايير الاختبارات والمقاييس.

فهذه المعايير ليست مطلقة أو مستقرة بل هى معايير نسبية تعتمد اعتمادًا أساسيًا على جماعات مرجعية معينة، وتتأثر هذه المعايير تأثرًا كبيرًا بتغير خصائص هذه الجماعات التى يقارن بها الفرد فى سمة معينة أو مجموعة من السمات التى يقيسها الاختبار.

ونظرًا للخصائص النسبية للمعايير وعدم استقرارها مما يحد من عمومية استخدامها فى تفسير درجات الاختبارات والمقاييس، فقد حاول علماء القياس التوصل إلى أسلوب جديد يهدف لجعل خصائص أدوات القياس مستقلة عن خصائص جماعات مرجعية معينة فى إطار نظرية معاصرة يطلق عليها نظرية الاستجابة للمفردة الاختبارية Item Response Theory (IRT)، وسوف نوضح أهم المفاهيم الأساسية لهذه النظرية وبعض تطبيقاتها التربوية والنفسية فى الباب الرابع.

محكات الأداء فى الاختبارات مرجعية المحك :

ذكرنا فى مستهل هذا الفصل أنه يمكن تفسير درجات الاختبارات تفسيرًا يمكننا من استخلاص معلومات مفيدة تتعلق بالأفراد بطرق متعددة. فإذا حولّنا الدرجات الخام إلى نوع آخر من الدرجات، مثل المئينيات أو التساعيات المعيارية أو الدرجات التائية، فإنه يمكننا مقارنة أداء الفرد بأداء جماعة مرجعية معيارية Normative Group. وهذ يرتبط بالقياس مرجعى الجماعة أو المعيار.



ولكن عندما نستخدم الاختبارات مرجعية المحك Criterion - Referenced Tests ، فإننا لا نهتم بالمركز النسبى للفرد بين أقرانه، وإنما نهتم بمقارنة درجته فى الاختبار بمستوى أداء Performance Standard يكون بمثابة محك Criterion يدل على المستوى المقبول لسلوك الفرد أو أدائه . فإذا كانت درجة الفرد تعادل أو تفوق مستوى الأداء المحدد فإن الفرد يكون قد حقق المستوى المطلوب، أما إذا قلت درجته عن المستوى المحدد فإنه أداءه يعد ضعيفًا.

فإذا لم تكن هناك قيود على عدد الأفراد الذين يمكن أن يحققوا مستوى الأداء المطلوب، فإن الاختبارات مرجعية المحك تكون مناسبة فى الكشف عن هؤلاء الأفراد. ولكن إذا اقتصرنا على انتقاء عدد محدود من الأفراد، وليكن مثلا نسبة ١٠٪ العليا منهم لتعيينهم فى إحدى الوظائف أو قبولهم فى إحدى الكليات فإن الاختبارات مرجعية الجماعة أو المعيار تكون أكثر ملاءمة فى هذه الحالة حيث يتم انتقاؤهم وفقًا لمعايير هذه الاختبارات التى أوضحناها فيما سبق.

فالإطار التفسيرى للاختبارات مرجعية المحك لا يعتمد إذن على جماعة مرجعية من الأفراد، وإنما يعتمد على مستوى أو محك أداء فى ضوء نطاق سلوكى Content Domain معين معرَّف تعريفًا جيدًا، يقيسه الاختبار . فمثلاً نحدد أداء الفرد فيما يتعلق بدرجة تمكُّنه من حل تمارين هندسية معينة، أو قراءة قطع نثرية محددة وهكذا.

فالتركيز فى هذه الحالة ينصب على ما يستطيع الفرد أداؤه، وما لا يستطيع أداؤه وليس على مقارنته بأقرانه دون تحديد لمعارفه أو مهاراته أو ما يجب أن يعرفه أو يؤديه . ويمكن تحديد محكات أو مستويات أداء أو درجات قطع للمهن والوظائف المختلفة للتحقق من كفايات العاملين بغض النظر عن مقارنة أداء أحدهم بأداء الآخر. كما يمكن تحديد محكات لمستويات كفاية الطلاب فى أوقات مختلفة أثناء عملية التعليم للحصول على معلومات تتعلق بأدائهم ومدى تقدمهم الدراسى.

ففى مثل هذه الحالات يكون هناك تناظر بين أداء الفرد ومحك الإنجاز المطلوب، وبذلك يمكن ترجمة درجة الفرد فى الاختبار مباشرة إلى سلوك محدد يمكن ملاحظته وقياسه. أى أن الدرجة الخام فى الاختبار يكون لها معنى فى ذاتها دون إسنادها إلى جماعة مرجعية من الأفراد. فإذا قلنا مثلاً أن الدرجة الخام ٢٣ التى حصل عليها طالب فى اختبار يشمل على ٢٥ مفردة اختيار من متعدد تقيس معرفة الرموز الكيميائية، فإن درجته تدل مباشرة على مدى معرفته بهذه الرموز. وإذا اخترنا الحد

الأدنى لمستوى الأداء المقبول ٩٠٪ ، فإنه يمكن تقييم أداء الطالب فى ضوء هذا المستوى المحدد. ونظرًا لأن الطالب أجاب عن نسبة ٩٢٪ من المفردات إجابة صحيحة، فإن أداءه يكون جيدًا، دون أن نحدد ما إذا كان أداؤه مرتفعًا أو منخفضًا بالنسبة لأقرانه الذين طبق عليهم الاختبار نفسه.

فمحكات الأداء في الاختبارات مرجعية المحك تعتمد على النسب المئوية Percentages من المفردات التى تقيس سلوكًا معرفًا تعريفًا دقيقًا والتى يجيب عنها الفرد إجابة صحيحة. بينما المعايير المئينية Percentiles للاختبارات مرجعية الجماعة أو المعيار تعتمد على النسب المئوية من الأفراد الذين تقل درجاتهم عن درجة معينة.

فالأولى تحدد ما يمتلكه الفرد من معارف ومهارات تتعلق بالأداء الجيد فى مهنة أو دراسة معينة، بينما تحدد الثانية ترتيب الفرد بين أقرانه فيما يقيسه الاختبار. فعلى الرغم من أن الطالب فى المثال السابق أجاب عن ٩٢٪ من مفردات الاختبار إجابة صحيحة مما يدل على أداء جيد، إلا أن درجاته ربما تقابل المئينى ٧٥ مثلاً، مما يعنى أنه يتفوق على ٧٥٪ فقط من أقرانه، وهذا يدل على ارتفاع مستوى الجماعة المرجعية التى استمدت منها المعايير المئينية للاختبار. أى أنه يمكننا تفسير درجات الاختبارات والمقاييس باستخدام النظامين المرجعيين المطلق Absolute والنسبى Relative معًا ، أى النظام مرجعى المحك والنظام مرجعى الجماعة،للحصول على معلومات متكاملة تتعلق بمستوى كفاية الفرد فى مجال معين ، وكذلك مكانته النسبية بين أقرانه فى هذا المجال.

لذلك فإن كلا من هذين النظامين يقدم معلومات مناسبة تفيد فى تقويم الأفراد والبرامج. واختيار أىٌ من النظامين ينبغى أن يعتمد على الغرض من عملية التقويم واستخدامات الاختبار. وسوف نوضح بمزيد من التفصيل طرق تحديد مستويات الأداء فى الاختبارات مرجعية المحك فى الفصل الثامن.

خلاصة

لكي يتمكن مستخدم الاختبار أو المقياس التربوى والنفسى من تفسير درجات الأفراد ينبغي أن يستند هذا التفسير إلى نظام مرجعي مناسب. فالاختبارات مرجعية الجماعة أو المعيار يستند تفسير درجاتها إلى معايير جماعة مرجعية، بينما يستند تفسير درجات الاختبارات مرجعية المحك إلى محكات أداء متوقَّع. والمعايير عبارة عن مجموعة من الدرجات المحوَّلة أو المشتقة من الدرجات الخام بطرق إحصائية معينة. وتوجد أنواع متعددة من معايير الاختبارات، مثل الدرجات المعيارية، والدرجات التائية، والمئينيات، والرتب المئينية، والتساعيات المعيارية، ونسبة الذكاء الانحرافية، ومعايير مكافئ العمر الزمني، ومعايير مكافئ الفرق الدراسية، والصفحات النفسية والتربوية.

وتختلف هذه المعايير فى طريقة تفسيرها للمركز النسبي للفرد فيما يقيسه الاختبار، وعلى خصائص وتكوين الجماعة المرجعية التى تستمد منها المعايير. أما محكات الأداء المتوقع فتدل على مستوى أداء مقبول لسلوك الفرد أو أدائه يمكن الحكم فى ضوئه على تحقق كفايات معينة لديه بغض النظر عن مقارنة أدائه بأداء أقرانه، أى أن تفسير الدرجات فى هذه الحالة يستند إلى نظام مرجعي مطلق وليس إلى نظام مرجعي نسبي كما فى الحالة الأولى. ومع هذا يمكن إحداث تكامل بين النظامين فى تفسير درجات بعض الاختبارات.

الفصل السابع

تحليل مفردات الاختبارات

* مقدمة
* أهمية تحليل مفردات الاختبار
* صعوبة مفردات الاختبار
* تصحيح قيمة معامل الصعوبة من أثر التخمين
* أوجه قصور معامل الصعوبة وكيفية تلافيها
* تمييز مفردات الاختبار
* أساليب تقييم «تمييز» المفردات استنادًا إلى محك داخلى
* أسلوب يعتمد على معامل الارتباط
* معامل تمييز المفردات
* استخدام معامل الارتباط الثنائى المتسلسل الحقيقى
* استخدام درجات المجموعتين الطرفيتين
* تقييم «تمييز» المفردات استنادًا إلى محك خارجى
* تفسير معامل تمييز المفردات وعلاقته بمعامل الصعوبة
* تحليل فاعلية بدائل الإجابات
* استخدام تحليل المفردات فى انتقاء مفردات الاختبار
* استخدام الحاسوب فى تحليل مفردات الاختبارات

مقدمة :

ناقشنا في الفصل الخامس الخصائص السيكومترية للاختبارات والمقاييس التربوية والنفسية، ومن أهمها الثبات والصدق، كما ناقشنا في الفصل السادس طرق تفسير درجات هذه الاختبارات والمقاييس باستخدام أنواع مختلفة من المعايير والمحكات. ولكن نظراً لأن الاختبار أو المقياس يشتمل على مجموعة من المفردات أو الفقرات أو الأسئلة، فإن خصائص الاختبار تعتمد أيضاً على خصائص مفرداته. ولكي نبني اختباراً يتميز بالفاعلية ينبغي فحص المفردات التي يشتمل عليها فحصاً دقيقاً باستخدام الأساليب الوصفية الكيفية، والأساليب الإحصائية الكمية. وقد سبق أن قدمنا بعض الأساليب الكيفية في تقدير صدق محتوى الاختبار، وسوف نتناول في هذا الفصل الأساليب الكمية في تحليل مفردات الاختبار Item Analysis ، وبخاصة الخصائص الإحصائية لكل مفردة من مفرداته.

وتتعلق هذه الخصائص بمستوى صعوبة المفردة Difficulty Level،ودرجة تمييزها Discrimination بين المستويات العليا والمستويات الدنيا في السمة أو القدرة التي يقيسها الاختبار في ضوء محك معين داخلي أو خارجي. فالمحك الداخلي ربما يتعلق بالأداء في مجموعة المفردات بعامة أو الاتساق الداخلي للمفردات، والمحك الخارجي يتعلق بالأداء خارج نطاق الاختبار نفسه مثل النجاح في دراسة معينة، أو في أداء عمل معين. والاهتمام في كلتا الحالتين ينصب على تقدير صدق كل مفردة من مفردات الاختبار.

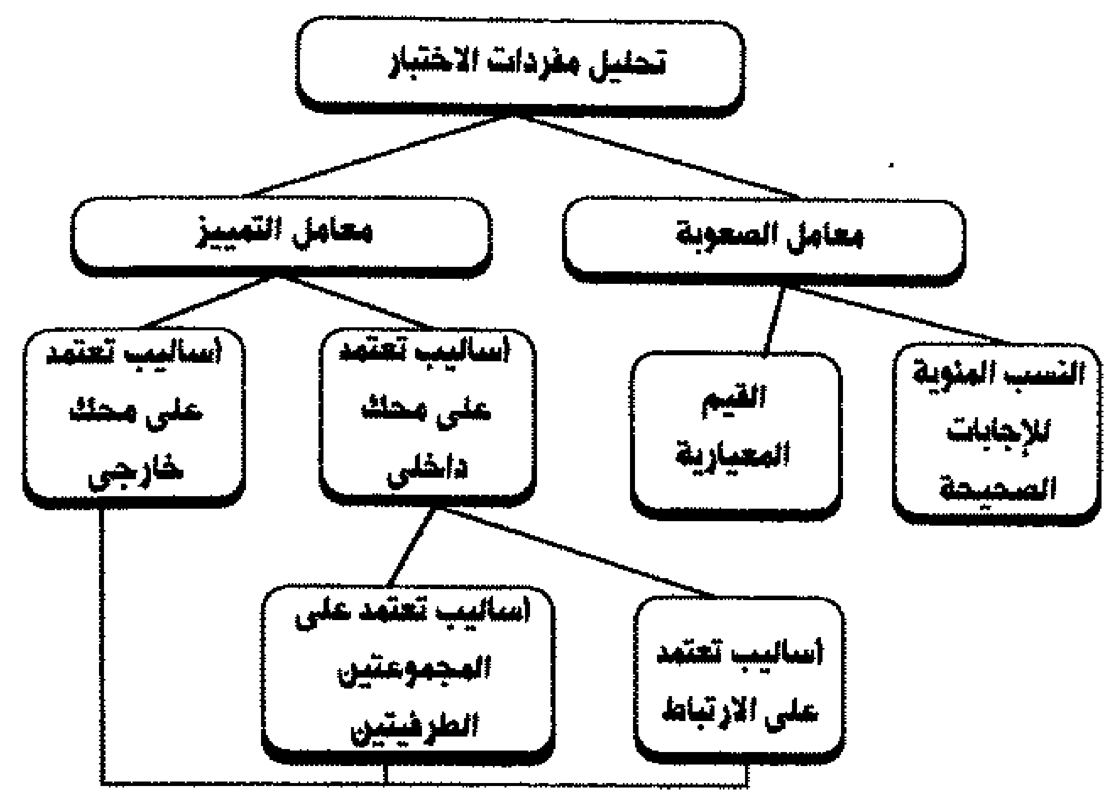
وسوف نناقش هاتين الخاصتين استناداً إلى أسس ومبادئ النظرية الكلاسيكية للاختبارات مرجعية الجماعة أو المعيار، ونرجئ مناقشتهما في ضوء نظرية الاستجابة للمفردات الاختبارية (IRT) إلى الفصل الرابع عشر. أما خصائص مفردات الاختبارات مرجعية المحك فسوف نوضحها في الفصل الثامن.

أهمية تحليل مفردات الاختبار:

يهتم القائمون ببناء الاختبارات والمقاييس بكتابة أو انتقاء مفردات عالية الجودة لقياس السمات الإنسانية قياسًا دقيقًا؛ لذلك يراعون كثيرًا من الشروط في تكوين هذه المفردات وصياغتها والتحقق بالأساليب المنطقية وأحكام الخبراء من صدق محتوى كل مفردة على حدة وكذلك صدق محتوى الاختبار ككل. غير أنه مهما بلغت دقة هذه الأساليب والأحكام فإنها لا تغنى عن التجريب الميدانى للاختبار وتحليل درجات مفرداته باستخدام الأساليب الإحصائية وتحديد العلاقة بين ما تقيسه المفردات وبين استجابات الأفراد لها، وذلك للتعرف على المفردات الغامضة أو المربكة أو التى تشجع على التخمين، والمفردات بالغة السهولة أو الصعوبة، حيث إنها لا تكشف عن الفروق الفردية فى السمة التى يقيسها الاختبار مرجعى الجماعة أو المعيار وبذلك لا تميز بين مستويات هذه السمة، كما أنها لا تسهم بأى قدر فى صدق الاختبار أو ثبات درجاته.

ويفيد ذلك فى المراجعة الفنية للمفردات وتحسينها بحيث تسهم كل مفردة إسهامًا إيجابيًا فيما يقيسه الاختبار. ويساعد تحليل المفردات أيضًا فى تنمية مهارات القائمين بإعداد الاختبارات والمقاييس نتيجة الفحص المستنير للمفردات، وتعرف جوانب الضعف التى ربما تجعل بعض هذه المفردات غير صالحة، واستبقاء المفردات التى تفى بخصائص تحقق الغرض من استخدام الاختبار أو المقياس فى مجالات التقويم المختلفة، وإعادة استخدامها إذا تطلب الأمر ذلك.

ولعل من المناسب قبل مناقشة أساليب تحليل مفردات الاختبار أن نقدم الشكل التخطيطى (٧- ١) التالى الذى يوضح كل من خاصتى الصعوبة والتمييز وأساليب تقدير قيمة كل منهما.



شكل (٧ ـ ١) يوضح أساليب تحليل مفردات الاختبار

(من تصميم المؤلف)

صعوبة مفردات الاختبار:

تعد صعوبة مفردات الاختبار من الخصائص التي تلعب دوراً مهماً في الاختبارات مرجعية الجماعة أو المعيار Norm - Referenced Tests وتؤثر في إجابات الأفراد عن مفرداتها . فالمفردات التي تشتمل عليها هذه الاختبارات ينبغي أن تميز تمييزاً دقيقاً بين مستويات السمة المراد قياسها، فالمفردة التي يجيب عنها جميع الأفراد، أو التي لا يستطيع أحدهم الإجابة عنها لا تفيد في الكشف عن الفروق بينهم فيما يقيسه الاختبار. وقد بينت الدراسات السيكومترية والإمبيريقية أن الاختبار يمكن أن يميز إلى أقصى حد ممكن بين الأفراد المختبرين إذا كان متوسط مستوى صعوبة المفردات التي يشتمل عليها ٠,٥٠ تقريباً، أي يستطيع أن يجيب ٥٠٪ منهم عن كل مفردة من مفرداته. غير أن التوزيع الأمثل لمستويات صعوبة المفردات في هذا النوع من الاختبارات يعتمد أيضاً على العلاقات القائمة بين مفردات الاختبار، فكلما زاد الارتباط بين المفردات تطلب الاختبار مدى متسع نسبياً لصعوبة مفرداته حول مستوى الصعوبة المتوسط ٠,٥٠ ، أما إذا قل الارتباط بين كل مفردة والأخرى فإننا نتوقع أن يكون مدى صعوبة المفردات أقل اتساعاً، لذلك يفضل انتقاء مفردات الاختبار بحيث يكون مدى صعوبتها معتدلاً، ولكن متوسط صعوبة جميع المفردات يكون ٠,٥٠ تقريباً.

ويمكن الحصول على قيم تشير إلى مستوى صعوبة مفردات الاختبار بإيجاد نسبة عدد الأفراد في جماعة مرجعية محددة الذين يجيبون عن كل مفردة إجابة صحيحة أو يستطيعون حل مشكلة معينة. وكلما زادت هذه النسبة دل ذلك على سهولة المفردة، وكلما قلت دل ذلك على صعوبة المفردة.

ويطلق على هذا المؤشر معامل صعوبة المفردة ($م_ص$) Item Difficulty Index.

$$\text{معامل صعوبة المفردة} = \frac{\text{عدد الأفراد الذين أجابوا إجابة صحيحة عن المفردة}}{\text{العدد الكلى لأفراد الجماعة المرجعية}}$$

$$\text{أى أن } م_ص = \frac{ح}{ن} \quad \text{.......... (٧ - ١)}$$

فإذا طبق اختبار يشتمل على ٥٠ مفردة اختيار من متعدد على مجموعة معينة من الأفراد، وأجاب ٣٠ فرد إجابة صحيحة عن أحد مفردات الاختبار، فإن معامل

$$\text{صعوبة هذه المفردة} = \frac{٣٠}{٥٠} = ٠,٦٠$$

وينبغي ملاحظة أن أقل قيمة لهذا المعامل (صفر) عندما تكون المفردة غاية فى الصعوبة ولم يجب عنها أحد إجابة صحيحة، وأكبر قيمة (١,٠٠) عندما تكون المفردة غاية فى السهولة، حيث يجيب عنها جميع الأفراد. ولذلك فإن القيمة ٠,٦٠ تعنى أن المفردة متوسطة الصعوبة إلى حد ما بالنسبة لجماعة مرجعية معينة.

وينبغى التأكيد هنا على أن خاصة «صعوبة المفردة» فى إطار النظرية الكلاسيكية للاختبارات ليست مقصورة على المفردة؛ لأنها تعكس أيضًا قدرة أفراد الجماعة المرجعية الذين أجابوا عن تلك المفردة.

لذلك يفضل عند تفسير قيمة معامل الصعوبة تحديد الجماعة المرجعية التى استندنا إليها فى تحديد هذه القيمة، فنقول مثلاً أنه إذا اختبرنا مجموعة ممثلة لتلاميذ نهاية مرحلة التعليم الأساسى بمحافظة معينة فى وقت ما بهذه المفردة، فإن قيمة معامل صعوبتها ٠,٦٠.

تصحيح قيمة معامل الصعوبة من أثر التخمين :

عند استخدام الصيغة (٧ ـ ١) في إيجاد معامل صعوبة مفردات الاختيار من متعدد، أو الصواب أو الخطأ التي تكون إجابتها إما صحيحة أو خطأ ينبغي مراعاة أثر التخمين Guessing .

فمن المعلوم أن بعض الأفراد يتوصلون إلى الإجابة الصحيحة عن طريق التخمين وحده أي بمحض الصدفة . فاحتمال التخمين الصحيح للإجابة عن مفردة الاختيار من متعدد التي تشتمل على خمسة بدائل ٠,٢٠ ، وأربعة بدائل ٠,٢٥ ، بينما احتمال التخمين الصحيح إذا كانت المفردة من نوع الصواب أو الخطأ ٠,٥٠ نظرًا لأنها تتضمن بديلين فقط مما يؤثر في قيمة معامل صعوبة المفردة.

لذلك يبرز التساؤل التالي : هل يمكن اعتبار الدرجة الكلية التي يحصل عليها فرد في الاختبار مساوية لعدد إجاباته الصحيحة عن مفردات الاختبار؟ أم أنه ينبغي تصحيحها من أثر التخمين العشوائي لاستبعاد ذلك الجزء من الدرجة الكلية الذي يفترض حصول الفرد عليه بالصدفة، أي دون أن يعرف بالفعل الإجابة الصحيحة عن بعض المفردات؟ والحقيقة أنه لا يوجد اتفاق بين علماء القياس حول إجابة هذا التساؤل.

فالبعض يرى أنه ليس من الضروري إجراء هذا التصحيح، وحجتهم في ذلك ما يلي :

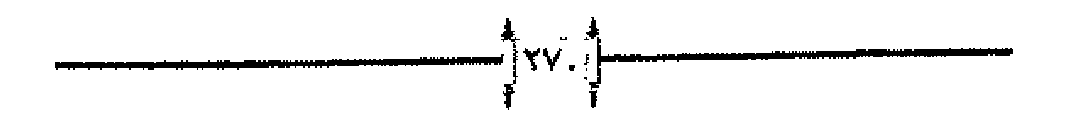
(١) لا يتغير ترتيب الأفراد فيما بينهم إذا أجرينا تصحيحًا لدرجاتهم الكلية في الاختبار من أثر التخمين.

(٢) يؤدي تصحيح درجات الأفراد الذين لا يلجأون إلى التخمين إلى خفض درجاتهم الكلية في الاختبار، بينما لا يحدث ذلك لمن اعتاد على التخمين في إجابته.

(٣) يعتمد تصحيح الدرجات من أثر التخمين على فرضية أن الإجابة الصحيحة إما تكون معلومة أو غير معلومة، في حين أن بعض الأفراد يلجأون إلى التخمين استنادًا إلى معلومات جزئية لديهم.

(٤) تصحيح الدرجات من أثر التخمين ربما يؤدي إلى الافتراض الخاطئ بأن الدرجات الناتجة هي الدرجات الحقيقية للأفراد في الاختبار.

لذلك يقترح أصحاب هذا التوجه بعض الإرشادات التي يمكن أن تسهم في تقليل أثر التخمين بقدر كبير كالتالي:



(١) تحذير المختبرين من التخمين العشوائى فى إجاباتهم عن مفردات الاختبار.

(٢) السماح بوقت كاف لكل فرد لكى يتمكن من محاولة إجابة كل مفردة من مفردات الاختبار.

(٣) تضمين البدائل التى تشتمل عليها مفردات الاختيار من متعدد معلومات صحيحة جزئيًا أو أخطاء شائعة تتعلق بالمجال الذى تقيسه المفردات.

أما البعض الآخر فيرى ضرورة تصحيح قيمة صعوبة مفردات الاختبار من أثر التخمين العشوائى؛ وذلك للحصول على تقدير نسبة عدد الأفراد الذين يجيبون إجابة صحيحة عن كل مفردة استنادًا إلى معلوماتهم الفعلية واستبعاد من يجيبون إجابة صحيحة عن طريق التخمين العشوائى.

ويستند هذا التقدير إلى افتراض أن اختيار الفرد إجابة غير صحيحة لمفردة اختبارية يرجع إلى افتقاره للمعلومات التى تتطلبها الإجابة الصحيحة ، وليس إلى خطأ معلوماته أو سوء فهمه، وكذلك افتراض أن الفرد الذى لا يعرف الإجابة الصحيحة يرى أن جميع البدائل التى تشتمل عليها المفردة يمكن أن تمثل الإجابة الصحيحة.

فإذا قبلنا هذين الافتراضين، فإننا نتوقع توصُّل نسبة معينة من الأفراد الذين لا يعرفون الإجابة الصحيحة إلى هذه الإجابة بمحض الصدفة، وهذه النسبة هى متوسط نسب عدد الأفراد الذين يختارون أحد البدائل الأخرى أى الإجابات الخطأ.

ويمكن الحصول على تقدير نسبة عدد الذين يجيبون إجابة صحيحة عن المفردة استنادًا إلى معلوماتهم الفعلية بطرح هذه النسبة من النسبة الكلية لعدد الأفراد الذين يجيبون عن المفردة إجابة صحيحة.

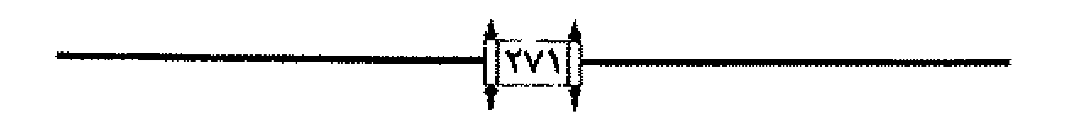
فلتصحيح درجة الفرد من أثر التخمين ($س_{ت}$) ، نستخدم الصيغة (٧ـ٢) التالية:

$$س_{ت} = ح - \frac{خ}{ك - ١} \qquad \text{.....(٧ـ٢)}$$

حيث (ح) ترمز إلى عدد الإجابات الصحيحة التى توصل إليها الفرد.

، (خ) ترمز إلى عدد إجاباته الخطأ.

، (ك) ترمز إلى عدد بدائل مفردة الاختيار من متعدد، أو مفردة الصواب أو الخطأ.



فإذا اشتمل اختبار على ٢٠ مفردة اختيار من متعدد لكل منها ٥ بدائل ، وأجاب فرد على ١٦ مفردة إجابة صحيحة ، فإنه يمكن تصحيح هذه الدرجة من أثر التخمين باستخدام الصيغة (٧ ـ ٢) كالتالى :

$$\text{س}_{\text{ت}} = ١٦ - \frac{٤}{٥ - ١} = ١٦ - ١ = ١٥$$

وإذا اشتمل اختبار على ٥٠ مفردة صواب أو خطأ ، وأجاب فرد عن ٣٨ مفردة إجابة صحيحة ، فإن درجته فى الاختبار بعد تصحيحها من أثر التخمين تصبح كالتالى :

$$\text{س}_{\text{ت}} = ٣٨ - \frac{١٢}{٢ - ١} = ٣٨ - ١٢ = ٢٦$$

وبذلك يمكن تقدير نسبة عدد الأفراد الذين يجيبون إجابة صحيحة عن مفردة الاختبار استناداً إلى معلوماتهم الفعلية ($\text{م}_{\text{ص}}$) باستخدام الصيغة (٧ ـ ٣) التالية :

$$\text{م}_{\text{ص}} = \frac{\text{ح} - \frac{\text{خ}}{\text{ك} - ١}}{\text{ن} - \text{نَ}} \quad .... (٧ - ٣)$$

حيث (ح) ترمز إلى عدد أفراد مجموعة المختبرين الذين أجابوا عن المفردة إجابة صحيحة .

، (خ) ترمز إلى عدد أفراد مجموعة المختبرين الذين أجابوا عن المفردة إجابة خطأ .

، (ك) ترمز إلى عدد بدائل المفردة .

، (ن) ترمز إلى العدد الكلى لأفراد مجموعة المختبرين .

، (نَ) ترمز إلى عدد أفراد مجموعة المختبرين الذين تركوا المفردة دون إجابة .

ويلاحظ أن هذه الصيغة أخذت بعين الاعتبار عدد الأفراد الذين لم يحاولوا الإجابة عن المفردة ، أى تركوها دون إجابة .

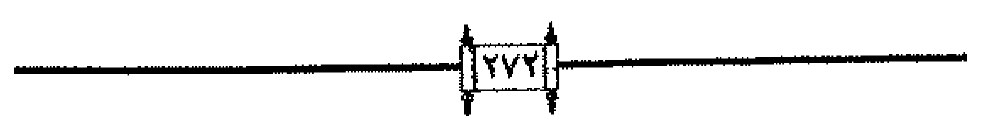
وإذا أردنا الحصول على قيمة معامل صعوبة مفردة دون إجراء تصحيح للدرجات من أثر التخمين ، فإننا نحذف النسبة $\frac{\text{خ}}{\text{ك} - ١}$ من البسط ، وتصبح ($\text{م}_{\text{ص}}$) كالتالى :

$$\text{م}_{\text{ص}} = \frac{\text{ح}}{\text{ن} - \text{نَ}} \quad .... (٧ - ٤)$$

فإذا لم يترك جميع الأفراد المختبرين أى مفردة دون إجابة ، فإن نَ = صفر ، وتصبح ($\text{م}_{\text{ص}}$) كالتالى :

$$\text{م}_{\text{ص}} = \frac{\text{ح}}{\text{ن}} \quad .... (٧ - ٥)$$

وهى نفس الصيغة (٧ ـ ١) التى ذكرناها من قبل .



ولتوضيح كيفية تطبيق الصيغة (٧ ـ ٣) نفترض أن الاختبار الذى يشتمل على ٢٠ مفردة اختيـار من متعدد ولكل منها ٥ بدائل طبق علـى ١٠٠ فرد حيث أجاب ٧٥ فرد منهم عن المفردة إجابة صحيحة، بينما ترك ١٠ أفراد المفردة دون إجابة، فعندئذ تكون صعوبة المفردة ($م_ص$) كالتالى:

$$م_ص = \frac{75 - \frac{15}{5-1}}{100 - 10} = \frac{75 - 3.75}{90}$$

$$= \frac{71.25}{90} = 0.79$$

أما إذا تغـاضينا عن تصحـيح الدرجات من أثر التـخمين، واستخدمنا الصـيغة (٧ـ٥) ، فإن صعوبة المفردة ($م_ص$) تكون كالتالى:

$$م_ص = \frac{75}{100} = 0.75$$

ويلاحظ أن المفردة تـبدو أقل صعوبة عندما استـخدمنا صيغة التـصحيح من أثر التخمين واستـبعدنا الأفراد الذين لم يحاولوا الإجابة عن المفـردة، وذلك لأننا اعتمدنا على نسبة عدد الأفراد الـذين أجابوا عن المفردة استناداً إلى معلوماتـهم الفعلية. أما إذا تغاضيـنا عن الأفراد الذين تركوا المـفردة دون إجابة ، فإن قيـمة معامل الصعـوبة بعد تصحيحه من أثر التخمين تصبح:

$$\frac{75 - 3.75}{100} = \frac{71.25}{100} = 0.71 \text{ تقريبًا.}$$

وهذا يعنى أن حـوالى ٧١٪ من المختـبرين أجابوا إجابـة صحيحة عن الـمفردة استناداً إلى معلوماتهم الفعلية ، فى حين أن ٧٥٪ مـن المختبرين أجابوا إجابة صحيحة عن المفردة سواء باستخدام معلوماتهم الفعلية أو عن طريق التخمين العشوائى.

غير أنه ينبغى مـراعاة أنه فى الاختبارات الموقوتة، أى التى تعتـمد على السرعة فى الإجابة ربما يترك بعض الأفراد إجابة بعض المفردات نظراً لضيق الوقت، ففى هذه الحالة لا نستطيع تـحديد ما إذا كانوا قد تركوا الإجابة عن هذه المـفردات بسبب ضيق الوقت أو لافتـقارهم إلى المعلومـات المطلوبة. فلكى نمـيز بينهما يمـكن افتراض أن جميع المفردات بما فى ذلك آخر مفردة أجاب عـنها الفرد قد حاول إجابتها استناداً إلى معلوماته، أما بقـية المفردات التى تركها دون إجابة فإننا نفـترض أنه لم يحاول الإجابة عنها.

كما ينبغى عند استخدام الصيغة (٧ ـ ٣) مراعاة الافتراضين اللذين ذكرناهما من قبل، على الرغم من أننا قد لا نستطيع عمليًا معرفة أيهما هو الذى أثر بالفعل فى إجابات الأفراد بحيث تعبر القيمة التقديرية لمعامل الصعوبة بعد تصحيحها من أثر التخمين عن نسبة عدد أفراد مجموعة المختبرين الذين توصلوا إلى الإجابة الصحيحة استنادًا إلى معلوماتهم الفعلية.

وعمومًا، فإن هذه القيمة التقديرية تعد قيمة تقريبية محكومة بالافتراضين اللذين استندت إليهما، ومع هذا فإنه من المتوقع أن تكون أكثر دقة من القيمة التى نحصل عليها دون إجراء هذا التصحيح.

أوجه قصور معامل الصعوبة وكيفية تلافيها :

إن استخدام إحدى الصيغ السابقة فى تقدير قيمة معامل صعوبة المفردات يتطلب مراعاة بعض الاعتبارات الأخرى نوجزها فيما يلى:

(١) ينبغى مراعاة أن قيم معامل صعوبة المفردات الاختبارية سواء أجرينا تصحيحها من أثر التخمين أو لم نجر ذلك تكون من المستوى الرتبى، فهذه القيم تدل على رتب الصعوبة النسبية للمفردات. فمثلاً إذا كانت قيم معامل صعوبة أربع مفردات ٠,٤٥ ، ٠,٤٠ ، ٠,٣٥ ، ٠,٣٠ على الترتيب ، فإن هذا يعنى أن المفردة الأولى أسهل المفردات الأربع، يليها المفردة الثانية، ثم الثالثة، وأصعبها المفردة الرابعة.

ولكننا لا نستطيع استنتاج أن الفرق بين القيمتين ٠,٤٥ ، ٠,٤٠ يساوى الفرق بين القيمتين ٠,٣٥ ، ٠,٣٠ ، أى أن هذه الفروق المتساوية بين قيم معامل صعوبة المفردات لا تناظر فروقًا متساوية فى مستوى صعوبة هذه المفردات، وهذا يعنى أن وحدات قيم معامل الصعوبة غير متساوية.

وهذا يعد أحد أوجه قصور معامل الصعوبة فى إطار النظرية الكلاسيكية للاختبارات . أما وجه القصور الآخر فهو اعتماد قيم معامل الصعوبة على خصائص الجماعة المرجعية التى استند إليها فى التوصل إلى هذه القيم مما يحد من إمكانية تعميم تفسيرها.

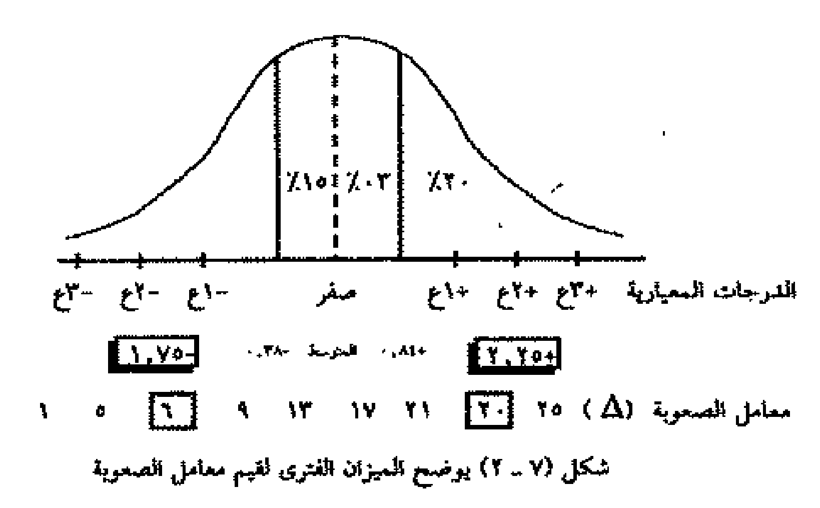
(٢) يرى ثورنديك (Thorndike, 1985) أنه يمكن التغلب على وجه القصور الأول بافتراض أن توزيع درجات السمة التى يقيسها الاختبار اعتدالى، وبذلك يمكن تحويل النسب المئوية من الأفراد الذين يجيبون عن مفردة اختبارية إجابة صحيحة



(أى القيم الرتبسية لمعامل الصعوبة) إلى الـميزان المتدرج ذى الفترات المـتساوية الذى يمثل المحور الأفقى للمنحنى الاعتدالى المعيارى (أى ميزان الدرجات المعيارية).

فقد سبق أن أوضحـنا فى الفصلين الثالث والرابع أنه يمكن تحـديد نسب مئوية معينة من المساحة تحت المنحنى الاعـتدالى تنحصر بين المتوسط ووحدات الانحراف المعيارى التى يمثلها المحور الأفقى.

فمثلاً المـساحة تحت المنحنى الاعتـدالى التى تنحصر بين المـتوسط وانحراف معيـارى واحد مُوجب أو سالـب تساوى ٣٤,١٤ ٪ وهكذا . فإذا كـانت قيمة مـعامل صعوبة مفردة ٠,٥٠ فإن هذه القيمة تناظر الدرجة المعيارية صفر على المحور الأفقى، وقيمة المعامل ٠,٦٥ مثلاً تناظر الدرجة المعيارية - ٠,٣٨ ، وقيمة المعامل ٠,٢٠ تناظر الدرجة المعيارية + ٠,٨٤ . ويمكن توضيح ذلك بشكل (٧ ـ ٢) التالى:



شكل (٧ ـ ٢) يوضح الميزان الفترى لقيم معامل الصعوبة

فقيمة معـامل الصعوبة ٠,٦٥ يعنى أن ٦٥٪ من أفراد الجماعة الـمعيارية أجابوا إجابة صحـيحة عن المفردة، وتـتضمن هذه النسبـة ٥٠٪ العليا مضافًا إلـيها ١٥٪ من الحالات التى تقل عن المتوسط كما هو موضح بالشكل.

لذلك فـإن هذه القيـمة تناظر - ٠,٣٨ وحـدة انحراف معـيارى على الـتدريج الأفقى. وبالمثل إذا كـانت قيمة معامل الصعوبـة ٠,١٦ فإن هذا يعنى أن ٢٠٪ أجابوا إجابة صحيـحة عن المفردة، أى ٥٠٪ مطروحًا منها ٣٠٪ مـن الحالات التى تزيد عن المتوسط فنجد أنها تناظر + ٠,٨٤ وحدة انحراف معيارى على التدريج الأفقى.

ويلاحظ أن القيم المرتفـعة لمعامل الصعوبة تناظر درجات معـيارية موجبة، أما القيم المنخفضة التى تدل على سهولة المفردات فإنها تناظر درجات معيارية سالبة.

كما يمكن إجراء بعض التحويلات الخطية المماثلة لتحويل الدرجات المعيارية إلى درجات تائية للتخلص من الإشارات السالبة. غير أن مركز الخدمات الاختبارية بالولايات المتحدة الأمريكية (ETS) قد توصل إلى نوع من التحويل الخطى يحقق ذلك، وأطلق على معامل الصعوبة الناتج الرمز (Δ) ، وتقرأ (دلتا) كالتالى:

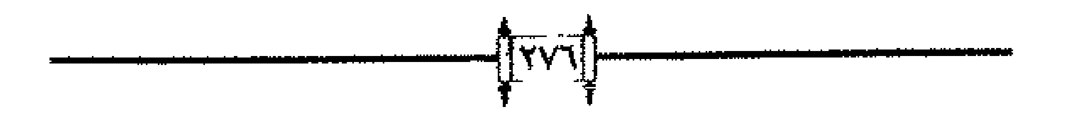
$$\Delta \quad = \quad ١٣ + ٤ \times د \quad .... \quad (٧ - ٦)$$

فالمفردة التى قيمة معامل صعوبتها صفر وحدة انحراف معيارى أى تقع عند المتوسط تكون قيمة (Δ) المناظرة لها ١٣. والمفردة التى قيمة معامل صعوبتها ـ ٣٨, ٠ تكون قيمة (Δ) المناظرة لها ١٣ + ٤ × (− ٣٨, ٠) = ١١,٥ ، أى ١٢ تقريبًا. وتتضمن الصيغة (٧ ـ ٦) ثابتين هما ١٣ ، ٤، أى أن قيمة متوسط الصعوبة (Δ) يساوى ١٣، وانحرافها المعيارى = ٤ . وهاتان القيمتان مناسبتان للتخلص من الإشارات السالبة، فمن الناحية العملية تكون أعلى قيمة لمعامل الصعوبة (Δ) مساوية ٢ وتناظر الدرجة المعيارية +٢,٢٥ ، وهذه تدل على مفردة صعبة؛ لأن نسبة من يجيب عنها إجابة صحيحة حوالى ١٢٪ ، وأقل قيمة لها تساوى ٦ حيث تناظر الدرجة المعيارية − ١,٧٥ ، وهذه تدل على مفردة سهلة ؛ لأن نسبة من يجيب عنها إجابة صحيحة حوالى ٩٦٪.

(٣) للتغلب على وجه القصور الآخر المتعلق بارتباط قيم معامل الصعوبة بخصائص الجماعة المرجعية حاولت النظرية السيكومترية المعاصرة التى أصبح يطلق عليها نظرية الاستجابة للمفردة الاختبارية (IRT) معالجة هذا القصور فى ضوء أسس ومبادئ جديدة .

والخلاصة أنه يمكن إيجاد قيم معامل صعوبة مفردات الاختيار من متعدد، أو الصواب أو الخطأ باستخدام إحدى الصيغ التى تعتمد على نسبة عدد الأفراد الذين يجيبون عن المفردة إجابة صحيحة، وميزان هذه النسب يكون رتبيًا مثل المئينيات، أو يمكن إيجاد هذه القيم على ميزان فترى باستخدام الدرجات المعيارية أو الدرجة المعيارية المعدلة (Δ).

وفى كلتا الحالتين ينبغى مراعاة الافتراضات التى تستند إليها كل منهما والتى ناقشناها فيما سبق.



تمييز مفردات الاختبار:

من الخصائص الأخرى المهمة التي ينبغي أن تتوافر في مفردات الاختبارات مرجعية الجماعة أو المعيار خاصة «التمييز Discrimination» ، ونعني بذلك مدى إمكانية قياس الفروق الفردية بواسطة مفردات هذه الاختبارات.

فإذا اختبرنا مجموعة من الطلاب باختبار تحصيلي مقنن في مجال دراسي معين، ووجدنا أن عدد الطلاب الضعفاء في المجال الدراسي الذي يقيسه الاختبار مساويًا عدد الطلاب الأقوياء الذين أجابوا إجابة صحيحة عن إحدى مفردات الاختبار، فإن هذه المفردة لا تصلح لقياس الفروق الفردية في التحصيل في هذا المجال الدراسي، حيث إنها لم تميز مطلقًا بين مجموعتي الطلاب.

غير أن هذا الاستنتاج يتطلب توفير محك نستند إليه في تحديد الطلاب الضعفاء والطلاب الأقوياء في التحصيل. فعلى الرغم من أن هناك محكات متعددة يمكن استخدامها في هذا الشأن، مثل آراء المعلمين وتقديرات الاختصاصيين في المجالات الدراسية المختلفة، إلا أن الدرجات الكلية التي يحصل عليها الطلاب في الاختبار نفسه تستخدم عادة كمحك مناسب سواء في الاختبارات التحصيلية المقننة أو الاختبارات الصفية التي يكتبها المعلم بنفسه لطلابه.

أي أن المحك المستخدم في هذه الاختبارات يكون محكًا داخليًا Internal Criterion . فإذا كان الطلاب الذين حصلوا على درجات مرتفعة في الاختبر قد أجابوا إجابة صحيحة عن إحدى مفردات الاختبار، والطلاب الذين حصلوا على درجات منخفضة في الاختبار قد أجابوا إجابة خطأ عن هذه المفردة، فإن المفردة تكون قد ميزت بين المجموعتين.

وربما نتخذ ذلك دليلاً على صدق المفردة Item Validity مع الأخذ في الاعتبار التحقق من صدق محتوى المفردة قبل تضمينها في الاختبار.

غير أنه في بناء الاختبارات التي تقيس الميول والصحة النفسية وبعض اختبارات الاستعدادات العقلية والجوانب المهارية يُفضل أن يكون المحك خارجي External Criterion ، فمثلاً في اختبار الميول يمكن استخدام أفراد المهن المختلفة كمجموعات محكة لتقدير صدق مفردات اختباراته الفرعية . فكل من هذه الاختبارات المتعلقة بالميل نحو مهنة معينة يشتمل على المفردات التي تميز بين الأفراد الذين ينتمون إلى هذه المهنة والناس بعامة. وقد سبق أن تناولنا في الفصل الخامس بمزيد من التفصيل أنواع المحكات وبعض المشكلات المتعلقة بها.



وسوف نوضح فيما يلي بعض الأساليب البسيطة التي يمكن استخدامها في تقييم مدى «تمييز» مفردات الاختبارات مرجعية الجماعة أو المعيار، وسوف نقدم بعض الأساليب الأخرى التي تناسب الاختبارات مرجعية المحك في الفصل الثامن.

أساليب تقييم «تمييز» المفردات استنادًا إلى محك داخلي:

تتعدد أساليب تقييم مدى تمييز مفردات الاختبارات استنادًا إلى محك داخلي، إذ يوجد ما يقرب من مائة أسلوب لتقدير علاقة درجات مفردة اختبار بالدرجة الكلية في الاختبار. غير أن معظم هذه الأساليب تعتمد على أنواع مختلفة من معاملات الارتباط لتحديد درجة هذه العلاقة، ويرجع اختلاف أنواع معاملات الارتباط المستخدمة إلى أسلوب معالجة الدرجات الكلية في الاختبار. فبعض الأساليب يعتبر هذه الدرجات متغيرًا متصلاً، وبعضها يعتبرها متغيرًا ثنائيًا وذلك بتقسيم الدرجة الكلية إلى قسمين متساويين، أو يعتمد على المقارنات الطرفية حيث يقسم الدرجة الكلية إلى ثلاثة أقسام ويقارن بين المجموعتين الطرفيتين **Extreme Groups**.

وتستند هذه الأساليب إلى افتراضين أساسيين متضمنين فيها كالتالي:

(١) تعبر الدرجة الكلية عما يقيسه الاختبار بالفعل، وتزداد جودة الاختبار إذا اشتمل على مفردات ترتبط درجاتها ارتباطًا مرتفعًا بالدرجة الكلية.

(٢) تعبر الدرجة الكلية عن سمة أحادية البعد، فإذا كان الاختبار يقيس سمات متعددة، فإنه يصعب تفسير معنى الدرجة الكلية، ويؤدي تحليل المفردات إلى نتائج غامضة.

ولتوضيح ذلك نفترض أن الاختبار يشتمل على عدد قليل من المفردات محتواها لفظي، وعدد كبير آخر من المفردات محتواها عددي، فإذا أجرينا تحليلاً لمفردات هذا الاختبار لتحديد علاقة درجات كل من هذه المفردات بالدرجات الكلية في الاختبار، فسوف تؤدي نتائج هذا التحليل إلى استبعاد المفردات التي محتواها لفظي؛ نظرًا لضعف علاقتها بالاختبار ككل، مما يقلل من صدق الاختبار. كذلك ينبغي في مثل هذه الحالة إجراء تحليل المفردات لكل محتوى على حدة، أي نوجد علاقة درجات كل مفردة محتواها عددي بالدرجات الكلية للمفردات التي محتواها عددي، وبالمثل في المفردات التي محتواها لفظي.

وسوف نقتصر هنا على أسلوبين فقط من أساليب تقييم مدى تمييز المفردات الاختبارية أحدهما يعتمد على الارتباط ، والآخر يعتمد على المقارنة بين مجموعتين طرفيتين نظرًا لشيوع استخدامهما وسهولة تطبيقهما وتفسير نتائجهما.

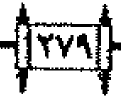
أسلوب يعتمد على معامل الارتباط :

يمكن تقييم تمييز مفردات الاختبار بإيجاد معامل الارتباط بين درجات كل مفردة والدرجات الكلية فى الاختبار، غير أن نوع معامل الارتباط المستخدم يعتمد على رؤيتنا لتوزيع الدرجات الكلية فى الاختبار أى المحك.

فإذا اعتبرنا توزيع هذه الدرجات متصلاً، فإنه يكون لدينا متغيران : أحدهما ثنائى الدرجة (درجات كل مفردة إما واحد صحيح أو صفر) ، والآخر فترى متصل (الدرجة الكلية فى الاختبار) ، وعندئذ يمكن استخدام معامل الارتباط الثنائى المتسلسل Biserial Correlation.غير أن إيجاد قيمة هذا المعامل تتطلب عمليات إحصائية تستغرق كثيرًا من الوقت والجهد.

لذلك ربما نلجأ إلى تقسيم المتغير المتصل. أى الدرجات الكلية فى الاختبار إلى قسمين متساويين يشتمل أحدهما على الدرجات العليا والآخر على الدرجات الدنيا، وعندئذ يكون لدينا متغيران أحدهما ثنائى الدرجة (درجات كل مفردة) ، والآخر فترى متصل (الدرجات الكلية فى الاختبار)، ولكننا قسمناه اعتباريًا إلى قسمين ، أى اعتبرناه أيضًا متغيرًا ثنائى الدرجة. وفى هذه الحالة يمكننا استخدام نوع آخر من معاملات الارتباط يسمى معامل الارتباط الثنائى المتسلسل الحقيقى Point Biserial Correlation .

ولكن هذا يعتمد أيضًا على ما إذا كنا نعتبر التقسيم الثنائى للمتغير المتصل تقسيمًا حقيقيًا أم مصطنعًا. فإذا اعتبرنا النجاح أو الفشل فى الإجابة عن مفردات الاختبار متغيرًا متصلاً يمثل السمة أو القدرة التى تنطوى عليها درجات المفردات، فإن نقطة تقسيم الدرجات الكلية تعتمد على مستوى صعوبة المفردات، وعندئذ يفضل استخدام معامل الارتباط الثنائى المتسلسل؛ نظرًا لأن قيم هذا المعامل لا تعتمد على نسبة عدد الدرجات العليا أو الدنيا، مما يجعل تقييم «تمييز» المفردات لا يتأثر بمستوى صعوبتها، وبذلك نحصل على تقديرين مستقلين لكل من التمييز والصعوبة.



أما إذا اعتبرنا أن النجاح أو الفشل فى الإجابة عن مفردات الاختبار يمثل كل منهما مجموعة متمايزة، وأن جميع أفراد المجموعة الأولى متكافئون فى النجاح، وجميع أفراد المجموعة الثانية متكافئون فى الرسوب، ففى هذه الحالة يتأثر مدى تمييز المفردات بمستوى صعوبتها.

لذلك فإنه فى حالة تساوى جميع الظروف الأخرى تكون قيمة معامل الارتباط الثنائى المتسلسل الحقيقى أقل من القيمة المناظرة إذا استخدمنا معامل الارتباط الثنائى المتسلسل.

ويرى بعض علماء القياس أن تقسيم الدرجات الكلية فى الاختبار إلى قسمين يعد تقسيمًا مصطنعًا وليس حقيقيًا، إذ إن هذه الدرجات تمثل مدى متصلا من السمة أو القدرة التى تقيسها مفردات الاختبار.

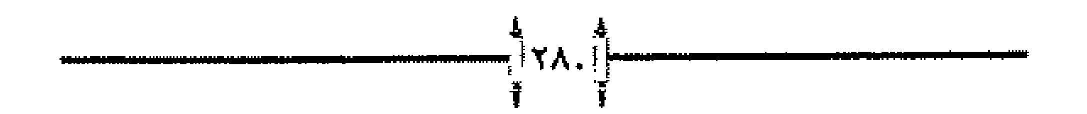
كما أن هذا التقسيم يؤدى إلى تأثر «تمييز» المفردات بمستوى «صعوبة» كل منها، وهو ما ينبغى تلافيه. لذلك ربما يكون معامل الارتباط الثنائى المتسلسل أكثر ملاءمة فى تقييم «تمييز» مفردات الاختبار، ومع هذا فإننا سوف نوضح كيفية استخدام معامل الارتباط المتسلسل الحقيقى فى تقييم «تمييز» مفردات الاختبارات نظرًا لسهولة حساب قيمته وشيوع استخدامه.

معامل تمييز المفردات:

أولا: استخدام معامل الارتباط الثنائى المتسلسل الحقيقى:

يمكن استخدام معامل الارتباط الثنائى المتسلسل الحقيقى Point Biserial Correlation كما ذكرنا فى إيجاد درجة العلاقة بين درجات مفردات الاختبار، والدرجات الكلية، وذلك بتقسيم توزيع الدرجات الكلية فى الاختبار إلى مجموعتين، إحداهما تمثل المجموعة العليا والأخرى تمثل المجموعة الدنيا فى السمة التى يقيسها الاختبار، وبذلك تكون نقطة التقسيم فى منتصف التوزيع.

والقيمة الناتجة تسمى: « معامل تمييز المفردات Discrimination Index»، وتدل هذه القيمة على مدى اتساق درجات كل مفردة مع درجات الاختبار ككل. ويمكن اشتقاق الصيغة الرياضية لمعامل الارتباط الثنائى المتسلسل الحقيقى($ر_{ث ح}$) من صيغة معامل ارتباط بيرسون (٣ ـ ١٠) التى قدمناها فى الفصل الثالث.



وهذه الصيغة كالتالى:

$$\text{ر}_{\text{ثح}} = \frac{\bar{\text{ص}}_{1} - \bar{\text{ص}}}{\text{ع}_{\text{ص}}} \sqrt{\frac{\text{ص}_{1}}{\text{ص}_{.}}} \quad \text{.........(٧ ـ ٧)}$$

حيث $\bar{\text{ص}}_1$ ترمز إلى متوسط توزيع الدرجات الكلية للمجموعة التى أجابت إجابة صحيحة عن مفردة معينة (أى س = ١).

، $\bar{\text{ص}}$ ترمز إلى متوسط توزيع الدرجات الكلية فى الاختبار.

، $\text{ع}_{\text{ص}}$ ترمز إلى الانحراف المعيارى للدرجات الكلية فى الاختبار.

، ص_1 ترمز إلى نسبة عدد أفراد المجموعة الأولى التى حصلت على (واحد صحيح) فى المفردة.

، $\text{ص}_{.}$ ترمز إلى نسبة عدد أفراد المجموعة الثانية التى حصلت على (صفر) فى المفردة.

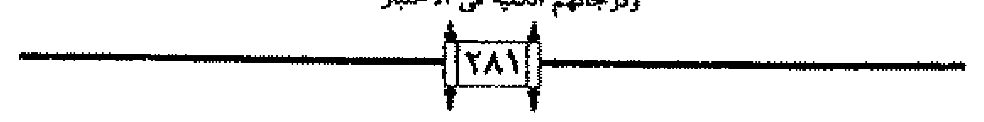
والمثال التالى يوضح كيفية إيجاد قيمة معامل تمييز مفردة اختبارية باستخدام هذه الصيغة:

نفترض أننا طبقنا اختباراً تحصيليا مقننا يشتمل على مفردات اختيار من متعدد على خمسة طلاب، وأردنا إيجاد قيمة معامل تمييز إحدى هذه المفردات .

ويوضح جدول (٧ ـ ١) التالى درجاتهم فى هذه المفردة (س) ، ودرجاتهم الكلية فى الاختبار (ص) ، وكذلك قيم (ص٢) ، وقيم (ص) عندما (س=١).

الطلاب	س	ص	ص٢	ص (عندما س=١)
١	صفر	٨	٦٤	-
٢	١	٦	٣٦	٦
٣	١	٩	٨١	٩
٤	صفر	٧	٤٩	-
٥	١	٥	٢٥	٥
المجموع	٣	٣٥	٢٥٥	٢٠

جدول (٧ ـ ١) يوضح درجات خمسة طلاب فى مفردة اختيار من متعدد ودرجاتهم الكلية فى الاختبار



فالخطوة الأولى : نوجد ($\bar{ص}_1$)، أى متوسط الدرجات الكلية فى الاختبار للمجموعة التى أجابت إجابة صحيحة عن المفردة (س = ١) من العمود الخامس كالتالى :

$$\bar{ص}_1 = \frac{٢٠}{٣} = ٦,٧ .$$

والخطوة الثانية: نوجد ($\bar{ص}$) أى متوسط الدرجات الكلية فى الاختبار من العمود الثانى كالتالى:

$$\bar{ص} = \frac{٣٥}{٥} = ٧ .$$

والخطوة الثالثة : نوجد ($ع_{ص_1}$) أى الانحراف المعيارى للدرجات من العمودين الثالث والرابع باستخدام الصيغة (٣ ـ ٤) التى سبق أن ذكرناها فى الفصل الثالث وهى :

$$ع_{ص} = \sqrt{\frac{مجـ (ص^٢) - \frac{(مجـ ص)^٢}{ن}}{ن}}$$

$$= \sqrt{\frac{٢٥٥ - \frac{(٣٥)^٢}{٥}}{٥}} = \sqrt{\frac{٢٥٥ - ٢٤٥}{٥}}$$

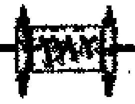
$$= \sqrt{\frac{١٠}{٢}} = \sqrt{٢} = ١,٤ \text{ تقريبًا}$$

والخطوة الرابعة : نوجد ($ص_{٠}$) أى نسبة عدد الطلاب الذين حصلوا على الدرجة (صفر) فى المفردة ، وذلك من العمود الثانى كالتالى:

$$ص_{٠} = \frac{٢}{٥} = ٠,٤٠ .$$

والخطوة الخامسة : نوجد ($ص_1$) أى نسبة عدد الطلاب الذين حصلوا على الدرجة (١) فى المفردة، وذلك من العمود الثانى أيضًا كالتالى:

$$ص_1 = \frac{٣}{٥} = ٠,٦٠ .$$



والخطوة السادسة : تطبق الصيغة (٧ ـ ٧) لإيجاد قيمة معامل الارتباط الثنائى المتسلسل كالتالى:

$$ر_{ثس} = \frac{\bar{ص}_1 - \bar{ص}_0}{ع_ص} \sqrt{\frac{ص_1}{ص_0}}$$

$$= \frac{٦,٧ - ٧}{١,٤} \sqrt{\frac{٠,٦}{٠,٤}}$$

$$= \frac{-٠,٣}{١,٤} \times ١,٣ = -٠,٢٨$$

ويلاحظ أن قيمة معامل الارتباط الثنائى المتسلسل لهذه المفردة سالبة، أى أن معامل تمييزها سالبًا .

وهذا يعنى أن نسبة عدد طلاب المجموعة الدنيا أى (الطلاب الضعاف) ولكنهم أجابوا إجابة صحيحة عن المفردة أكبر من نسبة عدد طلاب المجموعة العليا (الطلاب الأقوياء)، أى الذين حصلوا على درجات كلية مرتفعة فى الاختبار ، وأجابوا عنها إجابة صحيحة، وهذا ربما يدل على أن المفردة غامضة أو مضللة وينبغى حذفها من الاختبار أو مراجعتها من الناحية الفنية مراجعة دقيقة لتعرف وجه القصور فيها. وأحيانًا يحدث ذلك عن طريق الصدفة لقلة العدد الكلى للطلاب المختبرين كما فى هذا المثال.

ومن الجدير بالذكر أنه يمكن الحصول على هذه القيمة نفسها باستخدام صيغة معامل ارتباط بيرسون (٣ ـ ١٠) التى سبق أن ذكرناها فى الفصل الثالث، وذلك لأن معامل الارتباط الثنائى المتسلسل يعد حالة خاصة من معامل ارتباط بيرسون حيث يكون أحد المتغيران من النوع الثنائى.

وينبغى التنويه إلى أن هناك أساليب أخرى لإيجاد معامل التمييز تعتمد على أنواع أخرى من معاملات الارتباط مشتقة من معامل ارتباط بيرسون مثل معامل الارتباط الرباعى الحقيقى الذى يطلق عليه معامل فاى (ϕ) Phi Coefficient أو تعطى قيمًا تقديرية لمعامل ارتباط بيرسون مثل معامل الارتباط الرباعى Tetrachoric Correlation، وتعتمد جميعها على طريقة معالجة الدرجات الكلية فى الاختبار كمتغير متصل يتم



تقسيمه اعتباريًا إلى قسمين، أو يكون متغيراً ثنائيًا حقيقيًا، وقد أوضحنا هذه الأنواع من معاملات الارتباط فى مرجع آخر.

كما ينبغى التنويه أيضاً إلى أن عامل التخمين العشوائى يؤثر فى معامل التمييز مثلما يؤثر فى معامل الصعوبة. ويرى بعض علماء القياس أنه ينبغى تصحيح الدرجات من أثر التخمين قبل حساب قيمة معامل الارتباط الثنائى المتسلسل بين درجات المفردات والدرجات الكلية فى الاختبار أو غيره من المعاملات. غير أن الدراسات الإمبيريقية بينت أن تأثير استخدام الدرجات بعد تصحيحها على قيمة معامل الارتباط ضئيلاً عند حساب قيم معامل تمييز المفردات.

ثانياً : استخدام درجات المجموعتين الطرفيتين :

تعتمد هذه الطريقة على تقسيم الدرجات الكلية فى الاختبار إلى قسمين متمايزين، وبذلك نجعل المتغير المتصل Continuous Variable متغيراً ثنائياً Dichotomus Variable من أجل تبسيط العمليات الحسابية التى يتطلبها إيجاد معامل تمييز المفردات.

وتتميز هذه الطريقة بأنها لا تتطلب جهداً كبيراً، وتعطى قيمًا قريبة من القيم التى نحصل عليها باستخدام معامل الارتباط الثنائى المتسلسل الحقيقى.

وعادة يمثل أحد القسمين المجموعة التى نالت أعلى الدرجات فى الاختبار، ويمثل القسم الآخر المجموعة التى نالت أقل الدرجات فى الاختبار نفسه. وقد توصل كيلى Kelley إلى أفضل نسبة مئوية من الأفراد ينبغى أن تشتمل عليها كل من المجموعتين لكى يكون معامل التمييز أكثر دقة.

واستخدم فى ذلك أسلوب تحليل الانحدار للدرجات الكلية فى الاختبار فى تقدير الدرجات الحقيقية للأفراد، وأجرى بعض التحليلات العددية للدالة الجبرية الناتجة لمعرفة أقصى قيمة تصل إليها هذه الدالة، ووجد أن النسبة المئوية التى تقابل أقصى قيمة للدالة ٢٧,٠٢٦٧٨ .

لذلك أوصى كيلى Kelley عند تحليل مفردات الاختبار الاعتماد على النسبة ٢٧٪ من الأفراد فى كل من المجموعتين الطرفيتين ، واستبعاد نسبة ٤٦٪ الوسطى. ولكى تكون قيمة معامل التمييز متسقة من عينة إلى أخرى ينبغى عند إجراء هذا التقسيم ألا يقل عدد أفراد عينة تحليل مفردات الاختبارات المقننة عن ١٠٠ فى كل من المجموعتين الطرفيتين، ومن الأفضل زيادة عدد الأفراد عن ذلك.

أما فى حالة الاختبارات التحصيلية الصفية التى يكتبها المعلم لطلابه، فإن عدد طلاب الفصل يكون أقل من ذلك بكثير؛ لذلك يمكن ضم عدد من الفصول وتحليل

مفردات هذه الاختبارات استنادًا إلى العينة الكلية لكى تكون أكثر تمثيلاً وتؤدى إلى نتائج أكثر دقة.

وتتميز هذه الطريقة بأنها لا تتطلب جهدًا كبيرًا، وتعطى قيمًا قريبة من القيم التى نحصل عليها باستخدام معامل الارتباط الثنائى المتسلسل الحقيقى.

ويمكن إيجاد معامل تمييز مفردة اختبارية معينة بطرح نسبة عدد أفراد المجموعة الدنيا الذين أجابوا إجابة صحيحة عن المفردة (نِ)من نسبة عدد أفراد المجموعة العليا الذين أجابوا إجابة صحيحة عن هذه المفردة (نع)، أى أن :

معامل تمييز المفردة = نع − نِ ...... (٧ ـ ٨)

ويتطلب تطبيق الصيغة (٧ ـ ٨) اتباع ما يلى:

(١) تطبيق الاختبار على العينة المستهدفة، وتصحيحه، وإيجاد الدرجة الكلية التى حصل عليها كل فرد فى الاختبار.

(٢) ترتيب الدرجات الكلية من الأعلى إلى الأدنى للمجموعة ككل.

(٣) تقسيم الدرجات الكلية إلى قسمين بحيث يشتمل كل منهما على ٢٧٪ من عدد هذه الدرجات، ويمثل أحد القسمين المجموعة العليا، ويمثل الآخر المجموعة الدنيا. وإذا وقعت بعض الدرجات على الحد الفاصل بين المجموعتين، فإنه يمكن تعيين هذه الدرجات بطريقة عشوائية فى كل من المجموعتين بحيث يشتمل كل منهما على عدد متساو من الأفراد.

(٤) إيجاد عدد الأفراد الذين أجابوا عن كل بديل من بدائل مفردة الاختيار من متعدد فى كل من المجموعتين العليا والدنيا.

وهذا يتطلب عمل جدول مستطيل الشكل لكل مفردة، ويقسم إلى ثلاثة صفوف، يخصص الصف الأول للحروف المناظرة لبدائل الإجابات التى تشتمل عليها المفردة (أ، ب، جـ، د، هـ مثلاً)، ويخصص الصفان الثانى والثالث لعدد أفراد كل من المجموعتين العليا والدنيا الذين أجابوا عن كل من هذه البدائل على الترتيب.

وإذا كانت المفردة من نوع الصواب أو الخطأ، أو التكملة نوجد عدد الأفراد الذين أجابوا إجابة صحيحة عن المفردة فى كل من المجموعتين.

(٥) لإيجاد معامل تمييز المفردة نطرح نسبة عدد أفراد المجموعة الدنيا الذين أجابوا عن المفردة إجابة صحيحة من نسبة عدد أفراد المجموعة العليا الذين أجابوا عن المفردة إجابة صحيحة.

ويمكن توضيح هذه الخطوات بالمثال التالى حيث طبق اختبار يتكون من مفردات اختيار من متعدد لكل منها خمسة بدائل على ٤٠ طالب.

وفيما يلى درجاتهم الكلية فى الاختبار:

١٥	٢٠	٢٦	٤٣	٣٣	٢٣	٢٢	١٨
٢٩	٢٧	١٦	٣٥	١٧	(٤٥)	٢١	(١٣)
٢٧	١٩	٤٤	٤٢	٣٦	٣٧	٤١	٣٢
١٥	١٤	٣٥	١٧	٤٠	١٥	٣٢	٤٠
١٨	٤١	٣٠	٣٣	٤١	٣٤	٣٨	٣٧

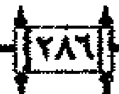
فالخطوة الأولى: نُرتب الدرجات ترتيبًا تنازليًا ، أى نبدأ بالدرجة ٤٥ وننتهى بالدرجة ١٣ .

والخطوة الثانية: نختار أعلى ٢٧٪ من الدرجات ، وأدنى ٢٧٪ من الدرجات لتمثل كل من المجموعتين العليا والدنيا على الترتيب، وذلك بضرب العدد الكلى للطلاب فى ٠,٢٧ ، أى: ٤٠ × ٠,٢٧ = ١٠,٨ ونقرب هذا العدد بحيث يصبح ١١ .

وبذلك تكون درجات المجموعة العليا هى : ٤٥، ٤٤، ٤٣، ٤٢، ٤١، ٤١، ٤١، ٤٠، ٤٠، ٣٨، ٣٧.

ودرجات المجموعة الدنيا هى: ١٩، ١٨، ١٨، ١٧، ١٧، ١٦، ١٥، ١٥، ١٥، ١٤، ١٣.

والخطوة الثالثة: نستخدم هاتان المجموعتان فى إيجاد قيمة معامل التمييز لإحدى مفردات الاختبار. وهذا يتطلب تحديد عدد طلاب كل من المجموعتين الذين أجابوا عن كل بديل من البدائل الخمسة التى تشتمل عليها المفردة، علمًا بأن البديل (جـ) يمثل الإجابة الصحيحة، لذلك أشرنا عليه بعلامة (*) ،وتلخيص ذلك فى جدول (٧ ـ ٢) التالى:



بدائل الإجابات	أ	ب	جـ*	د	هـ
عدد طلاب المجموعة الدنيا	١	صفر	٩	١	صفر
عدد طلاب المجموعة العليا	٤	٣	٢	١	١

جدول (٧ ـ ٢) يوضح عدد الطلاب الذين أجابوا عن

كل بديل من بدائل مفردة اختيار من متعدد

والخطوة الرابعة : نوجد قيمة معامل تمييز المفردة بطرح نسبة عدد طلاب المجموعة الدنيا الذين اختاروا البديل (جـ) من نسبة عدد طلاب المجموعة العليا الذين اختاروا هذا البديل حيث إنه يمثل الإجابة الصحيحة كالتالى:

معامل تمييز المفردة = $\frac{٩ - ٢}{١١}$ = $\frac{٧}{١١}$ = ٠,٦٤

وهذه القيمة موجبة مما يدل على أن المفردة تميز بين المجموعتين.

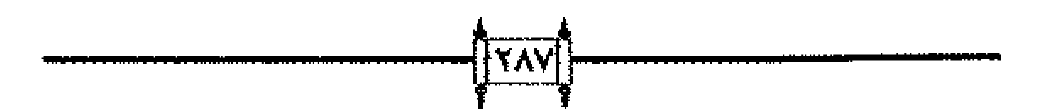
ومن الجدير بالذكر أن تفسير قيم معامل التمييز تفسيرًا أكثر دقة يتطلب مراعاة بعض الاعتبارات سوف نوضحها بعد قليل.

تقييم «تمييز» المفردات استنادًا إلى محك خارجي:

ذكرنا فيما سبق أن المحك الذى يستخدم فى كثير من الاختبارات وبخاصة الاختبارات التحصيلية يكون محكا داخليًا يتعلق بالدرجات الكلية فى الاختبار.

لذلك يرى ساكس Sax ونيتكو Nitko أنه إذا كان الغرض من الاختبار انتقاء الأفراد أو تصنيفهم، فإن كل مفردة من المفردات التى يشتمل عليها ينبغى أن تسهم فى الدرجة الكلية للاختبار، أى يكون معامل تمييزها موجبًا، كما ينبغى أن تسهم أيضًا فى درجات محك خارجى مستقل عن الاختبار.

ففى مثل هذه الحالة لا يجب أن نعرف فقط أن الأفراد الذين أجابوا إجابة صحيحة عن المفردة يتمايزون عن الذين أجابوا إجابة خطأ بالنسبة لدرجاتهم الكلية فى الاختبار، ولكن يجب أيضًا أن نعرف أن الأفراد الذين أجابوا إجابة صحيحة يحتمل بدرجة أكبر أن يحققوا نجاحًا فى محك خارجى، كأن يحصلوا على تقديرات مرتفعة من مشرفيهم على العمل أو من معلميهم. فمعامل التمييز الذى يعتمد على محك



الدرجات الكلية فى الاختبار، أى المحك الداخلى يدل على إسهام المفردات فى الاختبار ككل، ولكنه لا يدل على ما إذا كان الاختبار منبئًا جيدًا لمحك خارجى.

ويقترح Davis أنه فى التطبيقات الميدانية ينبغى انتقاء مفردات الاختبارات التى تُستخدم فى الانتقاء والتصنيف استنادًا إلى قيمتين لمعامل التمييز إحداهما تعتمد على ارتباط درجات المفردات بالدرجات الكلية فى الاختبار كمحك داخلى، والأخرى تعتمد على ارتباط الدرجات بالمتغير المراد التنبؤ به كمحك خارجى.

غير أنه فى الواقع ربما ترتبط بعض المفردات بالمحك الخارجى ولكنها لا ترتبط بالمحك الداخلى، والعكس ربما يكون أيضًا صحيحًا، هذا بالإضافة إلى المشكلات المتعلقة بالمحكات التى أوضحناها فى الفصل الخامس.

لذلك يرى نانلى (Nunnally,1978) أنه بدلا من بناء اختبار واحد يشتمل على مفردات تقيس جوانب مختلفة للمحك الخارجى، يُفضل بناء اختبارات منفصلة لكل جانب بحيث تكون مفردات كل منها مميزة تمييزًا مرتفعًا.

وعمومًا ، فإنه يندر استخدام محكات خارجية فى تقييم مدى تمييز مفردات الاختبارات، أى تقييم صدق المفردات، وإذا أردنا استخدامها فإنه يمكن الاسترشاد بنفس الطرق السابقة فى إيجاد قيم معامل التمييز.

تفسير معامل تمييز المفردات، وعلاقته بمعامل الصعوبة :

يعتمد تفسير معامل تمييز مفردات الاختبارات ومعامل صعوبتها على الغرض من استخدام الاختبار ، وخصائص الجماعة المرجعية التى تستمد منها هذه القيم.

وتتراوح قيم معامل التمييز بين -١ ، +١ ، فإذا كانت القيمة موجبة لمفردة معينة، فإن هذا يعنى أن تمييزها موجبًا، أى أن نسبة عدد الأفراد الذين حصلوا على درجات مرتفعة فى الاختبار ككل وأجابوا إجابة صحيحة عن هذه المفردة أكبر من نسبة عدد الأفراد الذين حصلوا على درجات منخفضة فى الاختبار وأجابوا عنها إجابة صحيحة أيضًا، والعكس إذا كانت القيمة سالبة. أما إذا كانت القيمة صفرًا، فإن المفردة لا تميز بين المجموعتين، ويمكن أن يحدث ذلك إذا كانت المفردة سهلة جدًا، أى أجاب عنها جميع الأفراد إجابة صحيحة، أو كانت صعبة جدًا بحيث لم يجب عنها أحد إجابة صحيحة، أو أن المفردة كانت غامضة.

وإذا كانت قيمة معامل تمييز المفردة سالبة، فإن هذا ربما يحدث عن طريق الصدفة لقلة عدد الأفراد المختبرين، أو ربما يرجع إلى غموض المفردة، أو عدم دقة الإجابة الصحيحة، أو قلة فاعلية بدائل الإجابات التى تشتمل عليها المفردة.

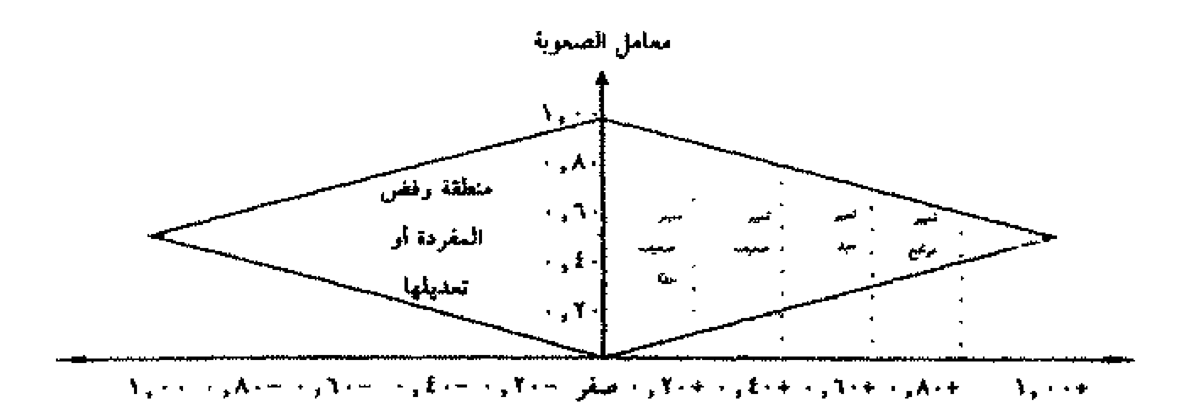
وبالطبع كلما زادت قيمة معامل تمييز المفردة كان ذلك أفضل، ولكن هذا يعتمد على الغرض من استخدام الاختبار كما ذكرنا. ففى الاختبارات مرجعية الجماعة أو المعيار التى تهدف للكشف عن الفروق الفردية ينبغى أن تميز مفرداتها تمييزاً مرتفعًا بين مستويات السمة التى يقيسها الاختبار. أما فى الاختبارات التشخيصية مرجعية المحك لا يجب أن نهتم كثيراً بمعامل تمييز مفرداتها، حيث يتركز الاهتمام هنا على جودة المفردات من حيث صياغتها، وقياسها لمكونات مهارات أو مهام معينة بحيث يمكن باستخدامها تحديد نقاط الضعف تحديداً دقيقًا.

وتعتمد قيم معامل تمييز المفردات على مستوى صعوبة المفردة، فكلما ابتعدت قيمة معامل صعوبة مفردة معينة عن ٠,٥٠ بالزيادة أو النقصان قلت القيمة القصوى لمعامل التمييز. لذلك لا يجب أن نتوقع قيمًا مرتفعة لمعامل التمييز إذا كانت مفردات الاختبار غاية فى السهولة أو الصعوبة.

وعمومًا ،إذا كانت قيمة معامل تمييز المفردة ٠,٤٠ أو أكثر، فإن هذا يكون دليلاً على أن المفردة تميز بدرجة جيدة بين المجموعتين الطرفيتين، وإذا تراوحت بين ٠,٢٠ ، ٠,٤٠ فإن تمييز المفردة بين المجموعتين يكون لا بأس به، وإذا قلت هذه القيمة عن ٠,٢٠ فإن تمييزها يكون ضعيفًا، أما إذا كان تمييزها صفراً أو سالبًا، فإنه ينبغى حذف هذه المفردة من الاختبار، أو مراجعتها مراجعة دقيقة وتعديلها.

ولكن ينبغى تأكيد أنه إذا كان الهدف من الاختبار قياس التمكن من مهارة معينة مثلاً، فإن قيمة معامل الصعوبة ربما تساوى ٠,٨٠ مثلاً، بينما قيمة معامل التمييز = صفر، ولا مانع من ذلك نظراً لأننا لا نهتم فى هذه الحالة بإبراز الفروق الفردية وإنما قياس درجة التمكن من المهارة المعينة.

ويوضح إيبل (Ebel, 1965) وكريسبن وفيلدهسن(Kryspin, Feldhusen, 1979) العلاقة بين قيم كل من معامل الصعوبة ومعامل التمييز للمفردات الاختبارية بالشكل التخطيطى (٧ ـ ٣) التالى:



معامل التمييز

شكل (٧ ـ ٣) يوضح العلاقة بين معامل الصعوبة

ومعامل التميز للمفردات الاختبارية

ويتضح من شكل (٧ ـ ٣) أن هناك علاقة بين قيم معامل صعوبة المفردات ومعامل تمييزها، فكلما ابتعدت قيمة معامل الصعوبة عن ٠,٥٠ إلى أعلى أو إلى أسفل على المحور الرأسى تقل القيمة القصوى لمعامل التمييز على المحور الأفقى.

لذلك فإن المفردات السهلة والمفردات الصعبة يكون تمييزها ضئيلاً، والمفردات التى تتراوح قيمة معامل صعوبتها بين ٠,٤٠ ، ٠,٦٠ يكون تمييزها مرتفعًا كما يوضحه الخطان الأفقيان، وإذا بلغت قيمة معامل الصعوبة صفراً أو واحداً صحيحًا، فإن معامل التمييز تكون قيمته صفراً فى الحالتين كما تُمثله نقطة تقاطع المحورين الأفقى والرأسى.

أما إذا كانت قيمة معامل تمييز المفردة سالبة ، أى وقعت فى منطقة الرفض التى إلى يسار المحور الرأسى، فإن هذه المفردة يمكن استبعادها أو مراجعتها وتدقيقها بالاسترشاد بنوع آخر من التحليل يسمى «تحليل فاعلية بدائل الإجابات» كما سيتضح بعد قليل.

فإذا عدنا إلى المثال الذى قدمناه والمتعلق بمفردة الاختيار من متعدد التى اشتملت على خمسة بدائل (جدول ٧ ـ ٢) حيث وجدنا أن قيمة معامل التمييز ٠,٦٤+ ،ومعامل صعوبتها دون تصحيح أثر التخمين ٠,٥٠ ، مما يدل على أنها مفردة متوسطة الصعوبة، وإذا عدنا إلى شكل (٧ ـ ٣) نجد أن قيمة معامل التمييز تدل على أن المفردة تميز تمييزاً مرتفعًا بين الطلاب الأقوياء والطلاب الضعاف فى التحصيل، وبذلك تعد مفردة جيدة.

تحليل فاعلية بدائل الإجابات :

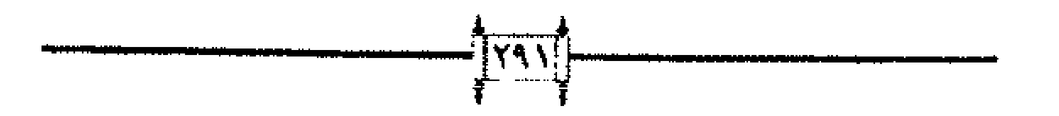
يهدف هذا التحليل لفحص البدائل الخطأ في مفردة اختيار من متعدد التي يطلق عليها المشتتات Distractors ، ويمكن إجراء هذا التحليل إذا كانت قيمة معامل تمييز المفردة سالبًا أو صفرًا أو قريبًا من الصفر كما يوضحه الشكل (٧ ـ ٣) . كما يمكن إجراء ذلك إذا لاحظنا أن المفردة تتطلب مزيدًا من التدقيق. فإذا تبين مثلاً أن أحد البدائل الخطأ (المشتتات) في مفردة معينة لم يختره أحد من الأفراد المختبرين، فإن هذا ربما يكون دليلاً على عدم فاعلية هذا المشتت أو وضوح خطأه، فمثل هذا المشتت لا فائدة منه لأنه يضيع وقت الفرد في قراءته دون أن يقيس ما يهدف الاختبار لقياسه.

وإذا وجدنا أن واحدًا من المشتتات اختاره معظم الأفراد وبخاصة الذين حصلوا على درجات مرتفعة في الاختبار ككل ، ولم يختاروا الإجابة الصحيحة، فإن هذا ربما يكون دليلاً على أن هذا المشتت أكثر جاذبية من الإجابة الصحيحة، أو ربما يكون هو الإجابة الصحيحة بالفعل، وليست الإجابة التي حددها من أعد الاختبار. وهنا ينبغي مراجعة الإجابة الصحيحة والتأكد من صحتها وعدم صحة هذا المشتت. فإذا تأكدنا أن هذا المشتت يمثل إجابة خطأ، فإنه ينبغي الإبقاء عليه؛ لأنه يساعد في كشف سوء الفهم وأنماط الأخطاء الشائعة لدى المختبرين مما يفيد في اقتراح أساليب العلاج.

أما إذا كانت إجابات الأفراد موزعة توزيعًا متعادلاً بين المشتتات الثلاثة أو الأربعة التي تشتمل عليها المفردة، ولكن المفردة غير مميزة، فإن هذا ربما يدل على أن هذه المفردة غامضة أو غاية في الصعوبة مما جعل المختبرين يلجأون إلى التخمين، وهكذا.

فإذا عدنا مرة أخرى إلى مثال مفردة الاختيار من متعدد (جدول ٧ ـ ٢) نجد أن المفردة تُميز تمييزًا موجبًا مرتفعًا بين المجموعتين العليا والدنيا حيث اختار الإجابة الصحيحة ٩ طلاب من المجموعة العليا، وطالبان فقط من المجموعة الدنيا. ولكن إذا نظرنا إلى المشتت (أ) نجد أن عدد طلاب المجموعة العليا الذين اختاروه (٤) أكبر من عدد طلاب المجموعة الدنيا (١) مما يدل على أن هذا المشتت يحتاج إلى إعادة نظر، وكذلك المشتت (ب)، والمشتت (هـ) حيث إنهما كانا أكثر جاذبية لطلاب المجموعة العليا. أما المشتت (د) فقد اختاره عدد متساو من كل من المجموعتين مما يدل أيضًا على عدم فاعليته.

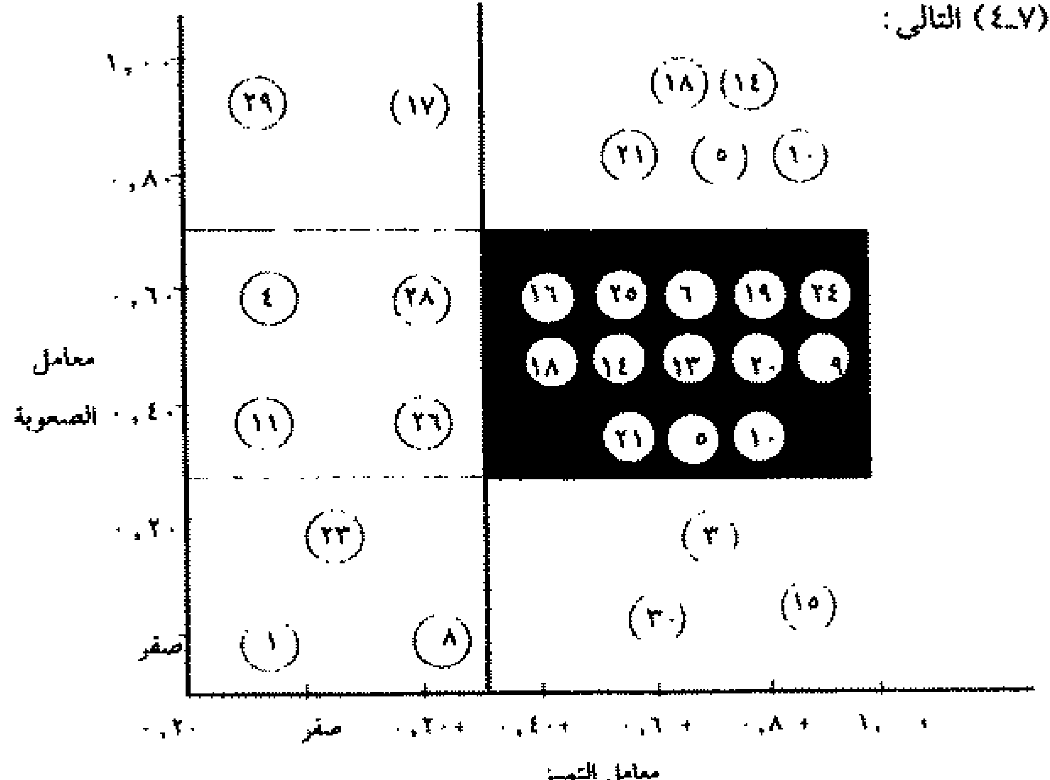
من هذا يتضح كيف أن تحليل فاعلية المشتتات لمفردات الاختيار من متعدد وغيرها من المفردات التي تتطلب الاختيار من بين بدائل معطاة يلقي الضوء على كل



وغيرها من المفردات التى تتطلب الاختيار من بين بدائل معطاة يلقى الضوء على كل مشتت تتضمنه المفردة، ويمكّن القائم ببناء الاختبار أو المعلم من فحص الإجابات فحصًا مدققًا يساعده فى تحسين المفردة وزيادة فاعليتها في قياس السمة التى يقيسها الاختبار، وتشخيص الأخطاء وسوء الفهم لدى المختبرين فى مجال معين.

استخدام تحليل المفردات في انتقاء مفردات الاختبار:

إن انتقاء المفردات فى الاختبارات مرجعية الجماعة أو المعيار يعتمد اعتمادًا كبيرًا على خصائص هذه المفردات، أى مستوى صعوبتها ودرجة تمييزها بين مستويات السمة المقاسة. ولذلك ينبغى عند انتقاء المفردات إجراء مقارنات فيما بينها بعد تجريبها ميدانيًا للتحقق من ملاءمة مستوى صعوبتها للأفراد المختبرين وتمييزها تمييزًا جيدًا بين الضعاف والأقوياء. ويمكن إجراء ذلك بتمثيل قيم كل من معامل التمييز ومعامل الصعوبة على المحورين الأفقى والرأسى على الترتيب. ويمكن أن تكون قيم معامل التمييز ناتجة عن استخدام معامل الارتباط الثنائي المتسلسل أو عن استخدام المجموعتين الطرفيتين، وتحديد نقطة تمثل قيمة كل من المعاملين لكل مفردة من مفردات اختبار يشتمل على ٣٠ مفردة اختيار من متعدد مثلا كما هو موضح بشكل (٧ـ٤) التالي:



شكل (٧ ـ٤) يوضح قيم معامل تمييز ومعامل صعوبة ٣ مفردة اختبارية

ويتضح من شكل (٧ ـ ٤) أن المنطقة المظللة تمثل المفردات التى تزيد قيم معامل تمييزها عن ٠,٣٠ ، وتنحصر قيم معامل صعوبتها بين ٠,٣٠ ، ٠,٧٠ فهذه المفردات يجب أن تستبقى فى الاختبار، أما المفردات التى تقع خارج هذه المنطقة فينبغى فحصها جيداً، نظراً لأن التحليل الإحصائى للمفردات لا يغنى عن فحص محتوى المفردة وطريقة صياغتها وتكوينها وإجراء التعديلات الضرورية عليها أو حذفها.

استخدام الحاسوب فى تحليل مفردات الاختبارات :

يتطلب تحليل مفردات الاختبارات التربوية والنفسية وتحليل الاختبارات ككل كما رأينا كثيراً من التحليلات الإحصائية والعمليات الحسابية التى ربما يصعب إجراؤها يدويًا أو باستخدام الآلة الحاسبة، وبخاصة إذا كان عدد الأفراد المختبرين كبيراً كما فى حالة الاختبارات المقننة. لذلك ينبغى الإفادة من برامج الحاسوب المتوافرة لإجراء كثير من التحليلات التى تتطلبها عمليات بناء الاختبارات والمقاييس.

فهناك برامج حديثة متعددة يمكن استخدامها فى تحليل مفردات الاختبار وفحصها وتجميع مفردات الاختبار وتخزينها فى ملف Item File يمكن استدعائه عند الحاجة وتعديل مفرداته وتحسينها وتنميتها.

ولعل برنامج MicroCAT يعد من البرامج الجيدة الشاملة، حيث يمكن باستخدامه إجراء تحليل مفردات الاختبارات فى إطار كل من النظرية الكلاسيكية للاختبارات (CTT) ، ونظرية الاستجابة للمفردة الاختبارية (IRT) ، وقد أعد هذا البرنامج مؤسسة نظم التقويم Assessment Systems Corporation (ASC) بالولايات المتحدة الأمريكية. وعلى الرغم من ارتفاع ثمن هذا البرنامج نسبيًا، إلا أنه يعد بمثابة نظام اختبارى متكامل حيث يطبِّق الاختبارات باستخدام الحاسوب، ويصحح نتائجها ويحلل درجات كل مفردة وكذلك درجات الاختبار ككل، ويعرض النتائج فى جداول منظمة ويمثلها برسوم وأشكال بيانية متنوعة موضحًا فيها أنماط استجابات كل طالب عن مفردات الاختبار.

كما أن هناك برنامجا Iteman لتحليل المفردات أعدته المؤسسة السابقة نفسها، ويعد البرنامج ملفات درجات المفردات اعتماداً على الفاحص الضوئى للاختبار Optical Test Scanner أو اعتماداً على إدخال البيانات يدويًا. ويمكن باستخدام هذا

البرنامج تحـليل درجات مفـردات الاختبار من متـعدد، والصـواب أو الخطأ، وموازين التقـدير، وما شابه ذلك. ويمكن إجـراء تحليل لمفردات الاخـتبارات التى تشـتمل على اختبارات فرعية قد يصل إلى عشرة اختبارات.

ويقدم البرنامج معلومات إحصائية تـتعلق بالمتوسط والتباين والانحراف المعيارى، والتواء وتفـرطح الدرجات الكلية فى كل اخـتبار، والوسيط وأعلى الدرجـات وأقلها، والتـوزيع التكرارى للدرجات، وقـيم معـامل كيـودر وريتشـاردسون لـثبـات درجات الاختبـارات، والخطأ المعيارى للقياس لكل اخـتبار فرعى، وكذلك قـيم كل من معامل صعوبة ومعامل تمييز المفردات للتعرف على المفردات التى تسهم فى قياس سمة معينة. ويمكن كذلك إجراء هذه التحليلات لمجموعات فـرعية من المختبرين كالذكور أو الإناث فقط. ويوضح شكل (٧ ـ ٥) التالى عينة من مخرجات هذا البرنامج:

شكل (٧ ـ ٥) يوضح عينة من مخرجات برنامج ITEMAN

لتحليل مفردات اختبار

ويمكن تحليل مفردات الاختبارات أيضًا اعتـمادًا على البرمجيات المتداولة بكثرة حيث يسهل استخدامها فى تصميم الجـداول الإلكترونية وتحليلها، ومن أهمها برنامج MICROSOFT EXCEL الذى يـمكن إدارتـه فـى بيئة النوافذ Windows 97 أو windows 95، وهذا يتطلب تنظيم الجداول وفق ما يراه مـعد الاختبار، وكذلك برمجة العمـليات الإحصائـية المطلوبة لـتحليل مفـردات الاختبار أو تـحليل درجاته الكـلية، وتصميم الرسوم ، والأشكال البيانية التى توضح النتائج توضيحًا بصريًا مناسبًا.

ويوضح شكل (٧ ـ ٦) التالى عينة من هذا التصميم أشار إليها وود ، وسافريت (Safrit &Wood, 1995) :

Microsoft Excel - IABOOK.XLS

File Edit Formula Format Data Options Macro Window Help

Item Analysis Spreadsheet

Student	Item 1	Item 2	Item 3	Item 4	Item 5	Total Score
1	1	1	1	1	1	5
2	1	1	1	0	1	4
3	1	1	1	0	1	4
4	1	1	1	1	0	4
5	1	1	1	1	0	4
6	1	0	1	0	1	3
7	1	0	1	0	1	3
8	0	1	1	0	1	3
9	1	0	0	0	1	2
10	0	1	0	1	0	2
11	0	1	0	0	1	2
12	0	1	1	0	0	2
Number Correct	8	9	9	4	8	
Difficulty	0.67	0.75	0.75	0.33	0.67	
Discrimination	0.66	0.29	0.68	0.42	0.12	
Decision	Retain	Revise	Revise	Revise	Discard	

شكل (٧ ـ ٦)

جدول إلكتروني لبرنامج EXCEL لتحليل المفردات

من هذا يتـضح أن تحليل مـفردات الاختبـارات والمقاييـس يُعد من العمـليات الأساسية فى جمـع البيانات الإمبيريقية المـتعلقة بكل مفردة وتقييم خـصائص كل منها

باستخدام قيم معامل الصعوبة ومعامل التمييز، واستخدام نتائج هذا التقييم فى فحص المفردات التى يتبين قلة فاعليتها ومراجعتها وتدقيقها أو حذفها إذا تطلب الأمر ذلك. كما يمكن استخدام هذه النتائج فى تقديم تغذية مرتدة للمختبرين والقائم بإعداد الاختبار، وبخاصة إذا كان معلماً حيث يمكنه الإفادة من النتائج فى إثراء تعلم طلابه وتنمية مهارته فى بناء الاختبارات.

غير أنه ينبغى إعادة تأكيد أن تحليل مفردات الاختبارات لا يغنى عن العناية والحرص فى بناء مفردات الاختبار وصياغتها صياغة دقيقة لا تحتمل الغموض أو التخمين. فانتقاء مفردات الاختبارات يعتمد بدرجة أساسية على الأحكام الإنسانية -بالإضافة إلى ما نتوصل إليه من نتائج التحليل الإحصائى للبيانات الإمبيريقية المتعلقة بالمفردات.

خلاصة

إن الغرض الأساسي من تحليل مفردات الاختبارات التحقق من بعض الخصائص التي ينبغي توافرها في المفردات التي يتكون منها الاختبار. ومن أهم هذه الخصائص صعوبة المفردة، ودرجة تمييزها بين المستويات العليا والدنيا في السمة أو القدرة التي يقيسها الاختبار. فمستوى صعوبة المفردة يتحدد بنسبة عدد الأفراد الذين أجابوا عن المفردة إجابة صحيحة، وكلما زادت قيمة هذه النسبة دل ذلك على سهولة المفردة، والعكس يكون صحيحًا . ولكن ينبغي مراعاة تصحيح هذه القيمة من أثر التخمين في المفردات التي تتطلب اختيار من متعدد، أو صواب أو خطأ. كما يمكن تلافي أوجه قصور معامل الصعوبة بأساليب متعددة كأن نجري على قيم هذا المعامل بعض التحويلات باستخدام خصائص المنحنى الاعتدالي، أو نستخدم أحد نماذج الاستجابة للمفردة الاختبارية التي سنوضحها في الفصل الرابع عشر. أما درجة تمييز المفردة فيمكن تقييمها استنادًا إلى محك داخلي باستخدام معامل الارتباط بين درجات كل مفردة والدرجات الكلية في الاختبار، أو باستخدام المجموعتين الطرفيتين، كما يمكن تقييمها استنادًا إلى محك خارجي حيث نوجد ارتباط درجات المفردات بدرجات المتغير الخارجي المراد التنبؤ به.

وينبغي مراعاة العلاقة بين معامل صعوبة المفردات، ومعامل تمييزها عند تفسير قيم كل منهما. ويمكن الإفادة من نتائج تحليل المفردات في التحقق من فاعلية بدائل الإجابات في مفردات الاختيار من متعدد، وتعرُّف جوانب القصور فيها من أجل مراجعتها وتعديلها أو استبعادها. ونظرًا لما يتطلبه هذا التحليل من عمليات إحصائية وحسابية متعددة، وبخاصة إذا كان عدد المفردات أو الأفراد أو كلاهما كبيرًا، فإنه يمكن إجراء ذلك باستخدام برامج الحاسوب مثل برنامج MicroCAT بطريقة أكثر مرونة وفاعلية.



تطبيقات القياس والتقويم التربوى والنفسى

(أولاً: قياس وتقويم الجوانب المعرفية



الفصل الثامن

قياس التحصيل والكفايات

* مقدمة
* قياس التحصيل الدراسى أو الأكاديمى من منظور تاريخى
* مفهوم التحصيل الدراسى وعلاقته بمفهوم الاستعداد
* التمييز بين اختبارات التحصيل واختبارات الاستعداد
* افتراضات أساسية يرتكن إليها قياس التحصيل
* تصنيف اختبارات التحصيل واستخداماتها

أولاً : اختبارات التحصيل مرجعية المحك

* خطوات بناء الاختبارات مرجعية المحك
* استخدامات الاختبارات مرجعية المحك

ثانيًا : اختبارات التحصيل مرجعية الجماعة أو المعيار

* خطوات بناء اختبارات التحصيل المقننة
* بعض أوجه النقد التى توجَّه للاختبارات التحصيلية المقننة
* خلاصة

مقدمة :

عرضنا في الباب الأول أهم الأسس المنطقية والمبادئ الأخلاقية والمفاهيم والأساليب الإحصائية والمنهجية للقياس والتقويم التربوى والنفسى.

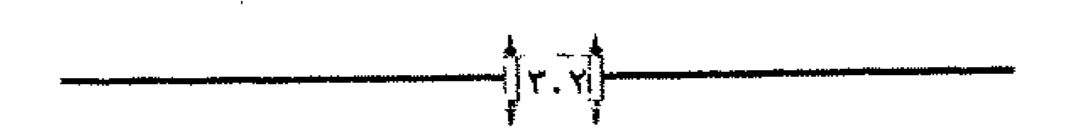
وسوف نتناول في هذا الباب بعض التطبيقات المتعلقة بقياس وتقويم الجوانب المعرفية للشخصية الإنسانية، وتوظيف البيانات المستمدة من الاختبارات التى تقيس هذه الجوانب في اتخاذ القرارات المتعلقة بالأفراد والبرامج. وتتضمن الجوانب المعرفية Cognitive Attributes التحصيل الدراسي أو الأكاديمي Academic Achievement، والذكاء العام General Intelligence، والاستعدادات الخاصة Special Aptitudes، حيث يتباين الأفراد تباينًا ملحوظًا في كل منها. فالفروق الفردية في التحصيل والذكاء والاستعدادات تبدو أكثر وضوحًا للمعلم في الصفوف المدرسية وبخاصة في مراحل التعليم العام.

ففي الصف المدرسي تكون الفرصة متاحة للمعلم لملاحظة طلابه في ظروف لا تتوافر مثلها في أى مكان آخر، فالمعلم يناقش طلابه في الصف ويتفاعل معهم ويضع تقديرات لأدائهم إما بواسطة اختبارات تحريرية أو شفوية، وبذلك يلمس الفروق الفردية في تحصيل طلابه بوضوح أثناء قيامه بعملية تعليمهم وتقويمهم. وهذا يجعل من الضرورى على المعلم تفهم طبيعة هذه الفروق وكيفية قياسها لكى يتمكن من الكشف عنها ومعرفة مداها، وبذلك يستطيع توجيه العملية التعليمية توجيهًا يناسب قدرات واستعدادات طلابه بما يمكنهم من اكتساب الكفايات المرجوة من العملية التعليمية.

وهذا يتطلب توضيح مفهوم التحصيل الدراسى أو الأكاديمي والعلاقة بينه وبين مفهوم الاستعداد والتمييز بين الاختبارات التى تقيس كل منهما.

لذلك سوف نفرد هذا الفصل لمناقشة كل من المفهومين، ونوضح أساليب قياس التحصيل الدراسى أو الأكاديمى باستخدام الاختبارات مرجعية الجماعة أو المعيار، والاختبارات مرجعية المحك، ونبين كيفية الإفادة من نتائج هذه الاختبارات في تقويم كفايات المتعلمين وفاعلية البرنامج التعليمى. ونتناول أساليب قياس الذكاء والاستعدادات الخاصة في الفصلين التاليين.

ولعل من المناسب أن نستهل هذا الفصل بإلقاء الضوء على تطور قياس التحصيل الدراسى أو الأكاديمى من منظور تاريخى.



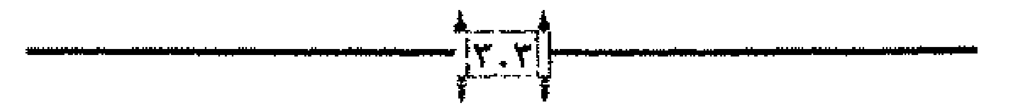
قياس التحصيل الدراسي أو الأكاديمي من منظور تاريخي :

إن استيعاب الأساليب المختلفة التي تستخدم في وقتنا الحاضر في قياس التحصيل الدراسي أو الأكاديمي استيعابًا مستنيرًا ينبغي أن يستند إلى رؤية واضحة للتطورات التاريخية التي مرت بها هذه الأساليب حتى أصبحت كما نراها اليوم.

فالامتحانات المدرسية كانت تعتمد اعتماداً أساسيًا على الاختبارات الشفوية حتى نهاية النصف الأول من القرن التاسع عشر. وفي عام ١٨٤٥ تزعم هوراس مان Horace Mann الذي كان يعد من القادة التربويين البارزين في ولاية بوسطون الأمريكية حركة تطوير التعليم العام، وأكد ضرورة استخدام الامتحانات التحريرية التي تتسم بالموضوعية والاتساق بدلاً من الاختبارات الشفوية في تقويم المتعلمين. ولكن استمر استخدام الاختبارات الشفوية في التقويم إلى أن بدأت في الاضمحلال وحلت محلها الاختبارات التحريرية. وفي عام ١٨٦٤ أعد المربي الإنجليزي جورج فيشر G. Fisher الذي كان يعمل مدير مدرسة أول اختبار تحصيلي تحريري يتكون من عدة مقاييس متدرجة وأمثلة ومواصفات متنوعة يمكن باستخدامها تقييم جودة الخط والنحو والتعبير والهجاء والرياضيات وغيرها من المواد الدراسية.

وأطلق على هذه البطارية من الاختبارات «كتاب الموازين Scale Book » . وقد وضح في هذا الكتاب كيفية تقدير درجات كل اختبار فرعي تشتمل عليه البطارية، والدرجات الكلية والمعايير التي يمكن استخدامها في مقارنة الطالب بنفسه في هذه المواد المختلفة، ورسم صفحة تربوية تبين جوانب قوته وضعفه. وبذلك وضع هذا الكتاب الاختباري الأساس لبناء الاختبارات التحريرية الموضوعية. ولعل هذه البطارية من الاختبارات كانت الركيزة التي استندت إليها حركة بناء الاختبارات مرجعية المحك في وقتنا الحاضر.

وبعد مضي ما يقرب من ربع قرن من الزمان أعد رايس J. Rice عام ١٨٩٥ أول اختبار تحصيلي موضوعي في الولايات المتحدة الأمريكية يقيس قدرة تلاميذ المدارس الابتدائية على الهجاء وطبقه على ما يقرب من ١٦٠٠٠ تلميذ في الصفوف من الرابع حتى الثامن، كما أعد اختباراً تحصيليًا في الحساب واللغة الإنجليزية. وتعد هذه الاختبارات لبنات بناء الاختبارات التحصيلية المقننة حيث شهد أوائل القرن العشرين ظهور عدد من هذه الاختبارات المقننة متمثلة في اختبار الحساب الذي أعده ستون Stone عام ١٩٠٨، واختبار جودة الخط الذي أعده ثورنديك Thorndike عام ١٩٠٩. وتزايد عدد هذه الاختبارات زيادة كبيرة قرب نهاية عام ١٩٢٠ بما في ذلك بطاريات الاختبارات، مثل بطارية اختبارات ستانفورد التحصيلية للمرحلة الابتدائية عام ١٩٢٣، وبطارية امتحانات أيوا للمحتوى الدراسي للمرحلة الثانوية عام ١٩٢٥.



وقد أسهمت هذه البطاريات وما أكدته كثير من الدراسات التى أجريت فى ذلك الوقت حول عدم اتساق تقديرات المعلمين لطلابهم، فى التوجه المتزايد نحو الاختبارات الموضوعية لتقويم التحصيل الدراسي. وتزايد عدد الاختبارات التحصيلية المقننة المنشورة التي استخدمت المفردات الموضوعية حتى وصل إلى المئات، وتضمنت اختبارات مادة دراسية واحدة، وبطاريات تشتمل على عدد من اختبارات مواد دراسية متعددة، واختبارات تشخيصية وتنبؤية. كما ازداد الاهتمام ببناء مفردات موضوعية تقيس الفهم والاستنباط وتطبيق المعلومات، وغيرها من العمليات المعرفية العليا، وكذلك الأهداف التعليمية الشاملة. وشهد عقد الثلاثينيات تصميم آلة تصحيح الاختبارات الموضوعية لتيسير تطبيق وتصحيح الاختبارات التحصيلية المقننة على نطاق واسع.

وعلى الرغم من الخلاف القائم حتى الآن حول فاعلية كل من الاختبارات الموضوعية التى تشتمل على مفردات اختيار من متعدد، أو صواب أو خطأ، أو مزاوجة، أو تكملة، أو اختبارات المقال، إلا أن تطوير هذه المفردات الموضوعية بحيث تشمل العمليات المعرفية العليا التى يُفترض أنها تقاس فقط بأسئلة المقال ساعد فى انتشار استخدامها فى الاختبارات التحصيلية المقننة لتقويم المتعلمين.

وقد أدى تصميم برامج اختبارية على المستوى القومى والإقليمى فى الولايات المتحدة الأمريكية إلى دفعة قوية لحركة تطوير الاختبارات التحصيلية وبخاصة الاختبارات المقننة، ولعل أهمها برنامج مكتب اختبارات القبول بالكليات (CEEB) عام ١٩٥٢الذى أشرنا إليه فى الفصل السادس، وبرنامج تقويم التعليم العام General Education Evaluation Program الذى اكتمل تنفيذه عام ١٩٥٤، وكذلك البرنامج الاختبارى للكليات الأمريكية American Colleges Testing Program عام ١٩٥٩ لفرز المتقدمين للالتحاق بالكليات بغرض انتقاء الطلاب المتميزين لإلحاقهم ببعثات دراسية.

ولم يقتصر استخدام الاختبارات التحصيلية المقننة على الأغراض التربوية، وإنما استُخدمت أيضًا فى انتقاء العاملين بالمصانع والمؤسسات الحكومية.

كذلك أدى التوسع فى استخدام هذه الاختبارات ـ بالإضافة إلى مشاركة خبراء علم النفس فى القياس النفسى وبناء الاختبارات المقننة ـ إلى التشابه الكبير فى عمليات بناء هذه الاختبارات وتقييمها سواء كانت تحصيلية أو استعدادات أو قدرة عقلية عامة. فالاختبارات التحصيلية المقننة التى تقيس أهدافًا تعليمية شاملة أو متسعة أصبحت تشبه اختبارات الذكاء العام، فالاختلاف بينهما يكمن فى مدى اتساع المحتوى وعموميته وما يتطلبه من توافر معارف أو مهارات تعليمية معينة لدى الأفراد المختبرين.

وعلى الرغم من انتشار استخدام الاختبارات التحصيلية المقننة ما يزيد عن نصف قرن فى مختلف مجالات التقويم التربوى والنفسى، إلا أنه ظهر فى أوائل الستينيات فى الولايات المتحدة الأمريكية كثير من الآراء المعارضة لهذا النوع من الاختبارات.

وحجتهم فى ذلك أن هذه الاختبارات تستند فى تفسير درجاتها على جماعة مرجعية معينة تعد بمثابة معيار يقارن فى ضوئه أداء الأفراد، لذلك أطلق عليها الاختبارات مرجعية الجماعة أو المعيار كما أشرنا فى الفصول السابقة، أى أنها تحدد المركز النسبى للفرد بين أقرانه، دون أن تهتم بتحديد ما يمكن للفرد أن يؤديه أو يفعله.

وعلى الرغم من أن ثورنديك Thorndike عام ١٩١٣، وفلاناجان Flanagan عام ١٩٥١ وغيرهما ميزا بين المعيار النسبى ومحك الأداء المطلق فى القياس التربوى والنفسى، إلا أن هذا التمييز لم يسهم إسهامًا ملحوظًا فى فلسفة بناء الاختبارات التحصيلية إلى أن نشر جيلزر Glaser مقالته المهمة حول الاختبارات مرجعية المحك فى دورية عالم النفس الأمريكى American Psychologist عام ١٩٦٣، وأعقبها مقالة بابام Popham وهاسك Husek فى دورية القياس التربوى عام ١٩٦٩ التى حفزت المشتغلين بالقياس على الاهتمام بتطوير أساليب التقويم باستخدام الاختبارات مرجعية المحك. وترتب على ذلك حركة بحثية كبيرة فى هذا المجال، حتى أنه بحلول عام ١٩٧٨ تم نشر ما يزيد عن ٦٠٠ دراسة وورقة تتعلق بهذا النوع الجديد من الاختبارات. واستمرت هذه الحركة البحثية ولا تزال حتى وقتنا الحاضر.

وقد شجع على ذلك الحركات التربوية التى نادت بالمسئولية التربوية Accountaibility، والتقويم بالأهداف Evaluation by Objectives وعقود الأداء Performance Contracting، والتعلم من أجل التمكن Mastery Learning، والتعلم من أجل التميز Learning for Excellence، وغيرها من التجديدات التربوية المعاصرة. غير أن كثيرًا من علماء القياس والمربين أصبحوا الآن يؤكدون أهمية إحداث تكامل بين نوعى الاختبارات، والإفادة منهما فى الأغراض المختلفة للتقويم التربوى والنفسى.

مفهوم التحصيل الدراسى وعلاقته بمفهوم الاستعداد:

يمكن تعريف التحصيل الدراسى Achievement بأنه درجة الاكتساب التى يحققها فرد أو مستوى النجاح الذى يحرزه أو يصل إليه فى مادة دراسية أو مجال تعليمى أو تدريبى معين. فالاختبارات التى يطبقها المعلم على طلابه على مدار العام

الدراسي مثل اختبار اللغة العربية أو الكيمياء أو الحاسوب يفترض أنها تقيس التحصيل الدراسي أو الأكاديمي.

والهدف من تصميم هذه الاختبارات التحصيلية هو قياس مدى استيعاب الطلاب لبعض المعارف والمفاهيم والمهارات المتعلقة بالمادة الدراسية في وقت معين أو في نهاية مدة تعليمية معينة.

أما الاستعداد Aptitude فيمكن تعريفه بأنه مدى قابلية الفرد Readiness للتعلم، أو مدى قدرته على اكتساب سلوك معين أو مهارة معينة إذا ما تهيأت له الظروف المناسبة. ومن هنا تبرز أهمية الاستعدادات لأنها تدل على إمكانية اكتساب أو تعلم الفرد لسلوك معين. وربما يختلف هذا السلوك المتعلم في درجة تعقده، فقد يكون مهارة عقلية معقدة مثل تعلم اللغات الأجنبية والرياضيات المتقدمة، أو تعلم أنشطة حركية أو جسمية بسيطة. ولذلك فإن تعريف الاستعداد يتضمن القدرة على تعلم مهارات متنوعة وسلوك متعدد. فالعبرة هنا بالقدرة على التعلم وليس نمط السلوك المتعلم أو نوع المهارة المكتسبة.

فالتحصيل الدراسي إذن يدل على الوضع الراهن لأداء الفرد أو ما تعلمه أو اكتسبه بالفعل من معارف ومهارات في برنامج تعليمي معين، أما الاستعداد فيدل على الأداء المستقبلي المتوقع. ولكن إذا تأملنا اختبارات التحصيل واختبارات الاستعداد نلاحظ أنها جميعًا تقيس ما تعلمه الفرد. وبعض اختبارات الاستعداد تقيس مجالات أكاديمية، أي تقيس تذكر المفاهيم والحقائق، وعمليات الاستنباط وحل المشكلات وغيرها من العمليات المعرفية التي تقيسها اختبارات التحصيل. لذلك ربما لا نستطيع التمييز بين اختبارين أحدهما يقيس التحصيل والآخر يقيس الاستعداد الدراسي مثلا بمجرد النظر إلى محتواهما. وأحيانًا تستخدم اختبارات التحصيل في التنبؤ بالأداء المستقبلي نظرًا لأن الأداء في وقت مضى أو الأداء الحالي يستخدم عادة في التنبؤ بدرجة جيدة بالأداء المستقبلي، وفي مثل هذه الحالة يمكن اعتبار اختبار التحصيل مقياسًا للاستعداد.

ويبدو من ذلك أن هناك تشابهًا بين التحصيل والاستعداد والاختبارات التي تقيس كل منهما. إلا أنهما يختلفان في مدى اتساع وعمومية الخبرات المتعلمة. فالاستعداد يعتمد على الخبرة التعليمية العامة، أي يعكس التأثير التجمعي للخبرات المتعددة التي يكتسبها الفرد في سياق حياته اليومية، أما التحصيل فيعتمد على خبرات تعليمية محددة في أحد المجالات الدراسية أو التدريبية.

التمييز بين اختبارات التحصيل واختبارات الاستعدادات :

يمكن التمييز بين اختبارات التحصيل واختبارات الاستعدادات على أساس التعريفات السابقة. فيمكن اعتبار اختبار ما مقياسًا للتحصيل إذا كان يقيس نواتج برنامج تعليمى أو تدريبى معين، واعتبار اختبار آخر مقياسًا للاستعداد إذا كان يقيس خبرات تعليمية واسعة، أى يعكس تأثير خبرات الفرد فى مجموعها التى اكتسبها فى حياته اليومية بهدف التنبؤ بما يستطيع الفرد أن يتعلمه فى المستقبل.

لذلك يمكن القول بأن الاختبارات التحصيلية تقيس التعلم الذى يتم تحت شروط محددة بدرجة نسبية وفى ظروف يمكن التحكم فيها، مثل التعلم الذى يتم داخل الصف المدرسى أو فى برنامج تدريبى معين، وأن التركيز يكون على الحاضر أو الماضى، أى ما تم تعلمه بالفعل.

أما اختبارات الاستعدادات مثل اختبارات الذكاء العام، وبطاريات الاستعدادات المتعددة، واختبارات الاستعدادات الخاصة، فهى تقيس تأثير التعلم تحت شروط يصعب التحكم فيها، وأن التركيز يكون على المستقبل، أو ما يمكن للفرد أداؤه فيما بعد إذا ما أتيحت له الفرصة المناسبة.

والتمييز الثانى بين اختبارات التحصيل واختبارات الاستعدادات يرتبط باستخدامات كل منهما، فاختبارات التحصيل تستخدم كما ذكرنا فى تقويم الفرد عند انتهائه من تعلم مادة دراسية معينة، أو موضوعًا دراسيًا معينًا، أو التدرب على مهارة معينة، وهى تبين ما يستطيع الفرد أداؤه فى وقت معين، أى أنها تركز على الحاضر أو الماضى وما تم تعلمه بالفعل. أما اختبارات الاستعدادات فتركز على المستقبل، أى تساعد فى التنبؤ بالأداء اللاحق، أو ما يمكن للفرد أداؤه مستقبلاً إذا ما أتيحت له الظروف المناسبة.

غير أن هذا التمييز غير قاطع، فقد يعتمد اختبار الاستعداد على ما تعلمه الفرد تحت شروط منتظمة إلى حد ما، وقد يشتمل اختبار التحصيل على خبرات تعليمية أو تدريبية شاملة وغير منتظمة، كما قد يستخدم كوسيلة تنبؤية للتعلم اللاحق. وفى مثل هذه الحالات ربما يُستخدم اختبار التحصيل لنفس الغرض الذى يُستخدم فيه اختبار الاستعداد.

فمثلاً ربما نستخدم درجة التلميذ فى اختبار تحصيلى فى الحساب للتنبؤ بنجاحه اللاحق فى الجبر، أو ربما نستخدم اختبارات التحصيل فى بعض المواد العلمية مثل الكيمياء والفيزياء والتاريخ الطبيعى للتنبؤ بأدائهم فى كليات الاقتصاد والعلوم السياسية، وهكذا.

وعند التمييز بين اختبارات التحصيل واختبارات الاستعدادات يجب أن يكون واضحًا أن جميع الاختبارات النفسية تقيس الوضع الراهن لسلوك الفرد، وتأثر درجة الفرد فى الاختبار بالخبرات السابقة لا يعنى بالطبع أنه لا يكون لها أثر لاحق أو مستقبلى فى أدائه.

وقد بين كرونباك Cronbach وأنستازى Anastasi أن علماء النفس اتجهوا إلى استخدام مفهوم جديد يحل محل مفهومى التحصيل والاستعداد هو مفهوم «القدرات النامية أو المستمرة Developing Abilities » .

فجميع اختبارات القدرات سواء كانت اختبارات ذكاء عام أو بطاريات استعدادات متعددة أو اختبارات استعدادات خاصة أو اختبارات تحصيل تقيس جميعها مستوى النمو الذى وصل إليه الفرد أو حققه فى قدرة واحدة أو أكثر. أى أنه يمكن اعتبار أن الاستعداد والتحصيل يقعان على متصل واحد Achievement - Aptitude Continuum ، ويتحدد موقع كل منهما على هذا المتصل وفقًا لنوعية وشمول الخبرات التى يفترض وجودها.

ويوضح شكل (٨ ـ ١) التالى ترتيب القدرات النامية على هذا المتصل:

نوعية ومحددة ـــــ خبـــــرات ـــــ عامة ومنتقلة الأثر

اختبارات تحصيل متعلقة بالمواد الدراسية (اختبارات صفية)	اختبارات تحصيل مقننة فى مادة دراسية واحدة	اختبارات تحصيل مقننة تقيس مهارات عقلية عامة	اختبارات ذكاء واستعدادات لفظية	اختبارات أداء (غير لفظية)	اختبارات غير متحيزة لثقافة معينة

شكل (٨ ـ ١) يوضح متصل اختبارات القدرات المعرفية

ويتضح من شكل (٨ ـ ١) أن اختبارات التحصيل المتعلقة بالمواد الدراسية وهى

الاختبارات الصفية التى يكتبها المعلم لطلابه لقياس معارف ومهارات محددة تحديدًا ضيقًا مثل اختبار قياس معانى الكلمات فى اللغات أو اختبار قياس المهارات الحسابية تقع فى أقصى الطرف الأيمن.

ويلى ذلك الاختبارات التحصيلية المقننة التى تستخدم فى قياس مدى تحقق أهداف تعليمية عريضة ومحتواها يكون متسعًا. وبعض هذه الاختبارات يتعلق بمادة دراسية واحدة، والبعض الآخر يقيس مهارات عقلية عامة مثل التفكير المنطقى وحل المشكلات، وهذه الاختبارات الأخيرة تشبه اختبارات الذكاء والاستعدادات. أما اختبارات الذكاء والاستعدادات اللفظية مثل اختبار ستانفورد بينيه Binet أو بطارية اختبارات الاستعدادات الفارقة (DAT) فإنها على الرغم من كونها مشابهة لاختبارات التحصيل، إلا أنها أقل ارتباطًا بمواد دراسية معينة. ويلى ذلك الاختبارات غير اللفظية التى تتطلب أداء عمليًا، ثم الاختبارات غير المتحيزة لثقافة معينة Culture Fair Tests والتى تستخدم لمقارنة أداء الأفراد من ذوى أطر ثقافية متباينة. وهذه الأنواع الأخيرة من الاختبارات لا ترتبط بمواد دراسية معينة، وإنما تتميز بالعمومية وإمكانية انتقال أثرها لمواقف حياتية مختلفة، ولذلك تقع فى أقصى الطرف الأيسر لمتصل الجوانب المعرفية.

وقد بينت الدراسات الإمبيريقية ارتباط الاختبارات المتجاورة على المتصل ارتباطًا مرتفعًا، ويقل الارتباط كلما ابتعدت عن بعضها على المتصل نفسه.

فمثلاً وجد أن قيم معامل الارتباط بين بعض اختبارات التحصيل والذكاء مرتفعة بدرجة كبيرة بعد أخذ أخطاء القياس فى الاعتبار.

افتراضات أساسية يرتكن إليها قياس التحصيل:

تصمم اختبارات التحصيل كما سبق أن ذكرنا لقياس المعارف والمهارات التى اكتسبها الطلاب فى المجالات الدراسية المختلفة، وكذلك لقياس مستويات المهارات فى المجالات التدريبية للعاملين فى المصانع والمؤسسات الحكومية والمجالات العسكرية وغيرها. ويتركز الاهتمام فى مثل هذه الحالات على قياس مكتسبات الأفراد فى نهاية مدة دراسية معينة يقدم خلالها مجموعة من الخبرات المنظمة.

لذلك يرتكن قياس التحصيل على بعض الافتراضات الأساسية لكى تكون نتائج القياس متسقة وصادقة وغير متحيزة وهى كالتالى:

(١) أن يقيس الاختبار نشاطًا سلوكيًا يمكن تحديده بدقة. فالاختبار التحصيلى سواء أعده المعلم لطلابه أو أعدته جهات أخرى ينبغى أن يستند إلى أهداف

تربوية يمكن صياغتها صياغة سلوكية محددة بحيث يمكن قياسها. وعلى الرغم من أن الاختبارات التحصيلية المقننة يكون محتواها أكثر اتساعًا من الاختبارات التى يعدها المعلم لطلابه إلا أنه ينبغى أن يكون هناك اتفاق على ما تقيسه من أهداف تشترك فيها مجموعة من المدارس أو المناطق التى سوف تستخدم الاختبار.

(٢) أن يقيس الاختبار الأهداف المتعلقة بالمادة الدراسية أو محتوى البرنامج التدريبى، ولا تقيس أهدافًا عارضة أو غير مهمة. وهذا يعنى أن يكون اختبار التحصيل صادقًا فى محتواه. فإذا صمم اختبار لقياس مهارة الطالب فى رسم الأشكال الهندسية مثلاً يجب أن يتضمن مفردات اختبارية تقيس هذه المهارة، ولا تقيس المهارة اللفظية.

(٣) أن تكون المعالجة التعليمية مناسبة للأفراد الذين سيطبق عليهم الاختبار بحيث تراعى الفروق الفردية بينهم، وأن تتاح لهم الفرصة لتعلم المجال الذى يقيسه الاختبار. وعلى الرغم من إمكانية توفير هذه الظروف فى الصفوف المدرسية التى يختبرها المعلم باختبارات من إعداده، إلا أن عدم تكافؤ هذه الفرص والخبرات التعليمية تعد مشكلة فى حالة اختبارات التحصيل المقننة نظرًا لتباين المحتوى التعليمى الذى تقيسه هذه الاختبارات من مدرسة إلى أخرى، أو من مؤسسة عمل إلى أخرى، واختلاف خبرات الأفراد المختبرين عن خبرات الجماعة المرجعية التى أعدت معايير الاختبار استنادًا إليها.

تصنيف اختبارات التحصيل واستخداماتها:

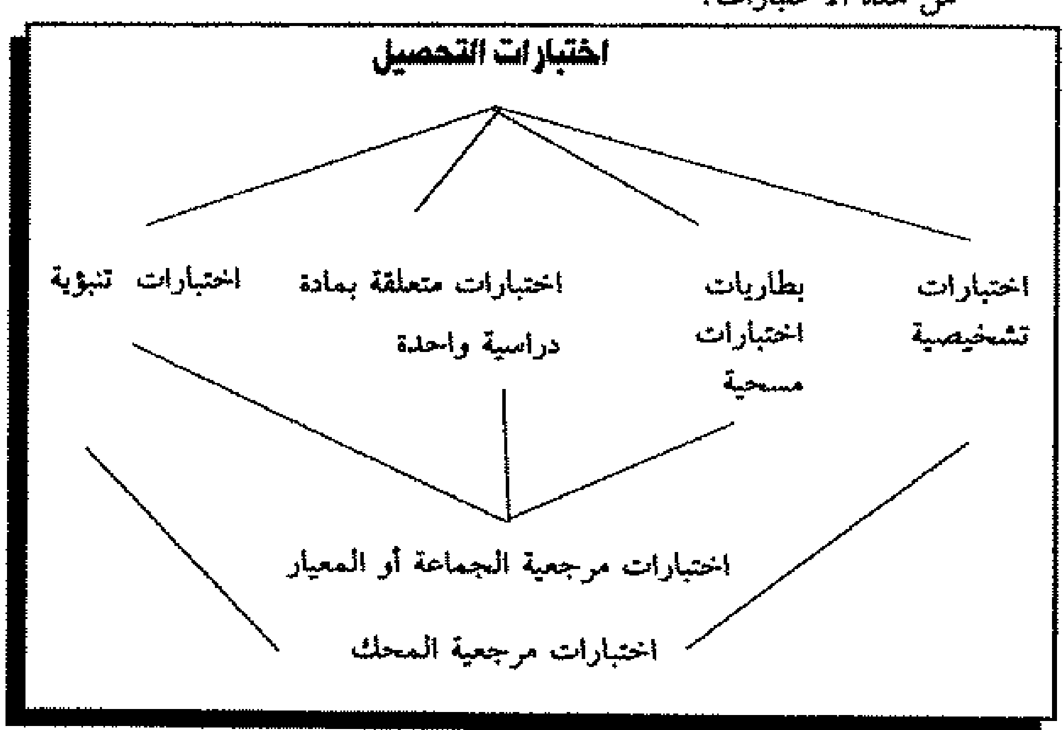
تتعدد أنواع اختبارات التحصيل وتتنوع استخداماتها فى المجالات التربوية والتدريسية. وعلى الرغم من إمكانية تصنيف هذه الاختبارات بطرق مختلفة، إلا أن التصنيف التالى يتضمن بعض الأبعاد الرئيسة التى يمكن فى ضوئها التمييز بين الأنواع المختلفة من هذه الاختبارات:

(١) **محتوى الاختبار:** على الرغم من أن معظم اختبارات التحصيل تقيس عينة من محتوى تعليمى أو تدريبى معين، إلا أن هذه الاختبارات تختلف فى مدى ضيق أو اتساع المحتوى، ومدى تمثيل عينة مفرداته لهذا المحتوى. فاختبارات التحصيل التى يُعدها المعلم لطلابه تكون محدودة فى محتواها حيث تقيس عادة وحدة تعليمية معينة فى إطار منهج دراسى محدد. إذ يمكن أن يقيس مثل هذا الاختبار معرفة الطلاب الصيغ الرياضية أو الرموز

الكيميائية ، أو يمكن أن يقيس مهارات الطلاب فى القراءة أو الجبر ، وهكذا.

أما الاختبارات التحصيلية المقننة فإن محتواها يكون أكثر اتساعًا ويشتمل على معلومات أساسية تتعلق بمادة دراسية واحدة لإحدى الفرق بغض النظر عن المنطقة التى تنتمى إليها المدرسة. كما أن بعض هذه الاختبارات يكون على شكل بطاريات Battery of Tests تقيس محتوى مواد دراسية مختلفة وليس مادة واحدة. وفى مثل هذه الحالة يُصنف الاختبار على أساس الفرقة أو المرحلة الدراسية، أى بطارية اختبار تحصيلى للمرحلة الأولى أو المرحلة الإعدادية أو الثانوية.

(٢) **الغرض من استخدام الاختبار:** تختلف اختبارات التحصيل تبعًا للأغراض التى تستخدم فيها . ويوضح الشكل التخطيطى (٨ ـ٢) التالى الأنواع الرئيسة من هذه الاختبارات:



شكل (٨ ـ ٢) يوضح الأنواع الرئيسة من اختبارات التحصيل

(من تصميم المؤلف)

ويتضح من شكل (٨ ـ ٢) أن هناك أربعة أنواع رئيسة من اختبارات التحصيل هى: الاختبارات التشخيصية Diagonstic Tests، وبطاريات الاختبارات المسحية

Survey Test Batteries ، والاختبارات المسحية المتعلقة بمحتوى مادة دراسية واحدة Single Survey Tests ، والاختبارات التنبؤية Prognostic Tests .

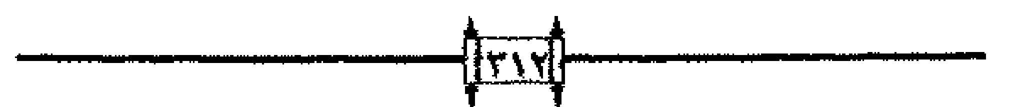
الاختبارات التشخيصية: تصمم هذه الاختبارات لتشخيص صعوبات تعلم الطلاب في مادة دراسية معينة، أي تهدف لتحديد جوانب القوة والضعف في تحصيل الطالب للمعارف والمهارات المتعلقة بهذه المادة.

ويتطلب بناء اختبار تشخيصي في المهارات الأساسية في القراءة أو الهجاء أو الحساب تحليل هذه المهارات إلى مكوناتها الفرعية، وقياس كل من هذه المكونات. لذلك فإن هذا النوع من الاختبارات يقيس التحصيل بدرجة أكثر تفصيلاً وعمقًا، وينقسم الاختبار عادة إلى عدد من الاختبارات الفرعية، وتقدر درجة لكل طالب في كل اختبار فرعي. فالاختبار التشخيصي في القراءة مثلا يستغرق وقتًا طويلاً نسبيًا في تطبيقه، ويشتمل على عدد كبير من المفردات المتنوعة. إذ ربما يتكون من عدد من الاختبارات الفرعية أحدها يقيس تعرف الكلمات، وآخر يقيس فهم معانى الكلمات، وثالث يقيس سرعة القراءة، ورابع يقيس فهم الموضوع المقروء، وغيرها، بل وربما يشتمل على أجهزة تقيس حركة العين، ومدى الانتباه، ووسع الذاكرة بواسطة عرض صور لمدة ثانية واحدة واختفائها، وتآزر اليد والعين، وهكذا. وتفيد هذه الاختبارات في التشخيص من أجل تقديم تعليم علاجي ، لذلك تكون عادة من نوع الاختبارات مرجعية المحك Criterion - Referenced Tests .

كما تفيد هذه الاختبارات المرشد النفسي والأخصائي الكلينيكي في تحديد صعوبات التعلم لدى بعض الأطفال، وتعرف أسبابها، واقتراح أساليب التدريب أو التعليم العلاجي المناسب. فالطفل البطئ في القراءة يحتاج إلى تدريب مناسب لزيادة سرعته، والطفل الذي يسلك في الصف المدرسي سلوكًا غير مرغوب فيه بسبب ضعف تحصيله أو تكرار رسوبه، والطفل المتفوق في التحصيل ربما يعاني أيضًا من الملل مما يؤدي به إلى سوء التوافق وربما الهروب من المدرسة، فكلاهما يحتاج إلى تعديل سلوكه بأساليب علاجية منظمة استنادًا إلى نتائج الاختبارات التشخيصية.

بطاريات الاختبارات المسحية: تتكون كل منها من مجموعة من الاختبارات المسحية يتعلق كل منها بمادة دراسية مختلفة لإحدى فرق مرحلة تعليمية معينة. فمثلا بطارية الاختبارات المسحية لإحدى فرق المرحلة الابتدائية ربما تشتمل على اختبارات في الحساب، والقراءة، والدراسات الاجتماعية، والعلوم، وربما الهجاء.

وبطارية الاختبارات المسحية للمرحلة الثانوية ربما تشتمل على اختبارات في اللغة العربية، والعلوم، والمواد الاجتماعية. وهذه البطاريات عبارة عن مجموعة متناسقة من اختبارات التحصيل المسحية التي يمكن تطبيق كل منها على حدة،



وتُستخدم فى تحديد المستوى العام لأداء الطالب فى المواد الدراسية ومقارنة أدائه بأداء أقرانه ، لذلك تعد من نوع الاختبارات مرجعية الجماعة أو المـ [illegible] Norm Referenced Tests . ويمكن أيضًا تمثيل درجات الطالب فى صفحة تربوية Profile للمقارنة بين أدائه فى مختلف المواد الدراسية التى تقيسها بطارية الاختبارات بعد تحويل الدرجات الخام إلى درجات معيارية ذات تداريج موحدة على الرغم من اختلاف عدد مفردات كل اختبار تشتمل عليه البطارية كما سبق أن أوضحنا فى الفصل السادس. ومثال ذلك بطارية كاليفورنيا للاختبارات المسحية California Achievement Tests Battery (CAT) ، وبطارية متابعة التقدم التربوى للطلاب من الصف الرابع حتى نهاية مرحلة التعليم الثانوى Sequential Tests of Educational Progress (STEP)، وبطارية ستانفورد لاختبارات التحصيل لجميع صفوف التعليم العام Stanford Achievement Tests Battery ، وبطارية ستانفورد للمهارات الأكاديمية Stanford Tests of Academic Skills (TASK)، وتشتمل على مستويين أحدهما للمرحلة المتوسطة والآخر للمرحلة الثانوية ، وتقيس مهارات القراءة، وفهم المقروء، واللغة الإنجليزية، والرياضيات.

وتختلف بطاريات اختبار التحصيل المسحية عن الاختبارات التشخيصية فى أن محتواها يكون متسعًا، ويشتمل كل اختبار على عينة محدودة من المحتوى والمهارات فى المواد الدراسية المختلفة، لذلك يصعب باستخدامها تشخيص جوانب القوة والضعف فى تحصيل الطلاب.

ولعل تطبيق بطارية الاختبارات المسحية لتحديد المستوى العام لأداء الطالب يعد الخطوة الأولى المنطقية، حيث يليها تطبيق الاختبارات التشخيصية المتعلقة بالمجال أو المهارة التى يفتقر إليها الطالب لكى يتسنى التحديد الدقيق لجوانب القصور فى هذا المجال أو هذه المهارة واقتراح العلاج المناسب.

الاختبارات المسحية المتعلقة بمادة دراسية واحدة : وتهدف هذه الاختبارات المسحية إلى قياس التحصيل فى مادة دراسية واحدة. فعلى الرغم من إمكانية تطبيق أحد اختبارات بطاريات الاختبارات المسحية التى أشرنا إليها بمفردها، إلا أنه يفضَّل استخدام الاختبارات المتعلقة بمادة دراسية معينة؛ وذلك لأن عينة محتواها تكون أكثر تمثيلاً لهذه المادة، وعدد مفردات الاختبار يكون أكبر مما يسمح بتقويم التحصيل بدرجة أكثر تفصيلاً. غير أن هذه الاختبارات تُزوِّدنا بدرجة كلية واحدة فقط بدلاً من عدة درجات كما فى بطاريات الاختبارات المسحية، وبذلك لا تساعد المعلم فى تعرف مجال ضعف الطالب فى المادة الدراسية.

الاختبارات التنبؤية : على الرغم من أن اختبارات التحصيل تقيس الوضع الراهن لأداء الطالب فى مادة دراسية معينة ، إلا أنه يمكن استخدام هذه الاختبارات فى التنبؤ بالتحصيل اللاحق فى هذه المادة، وبذلك تؤدى وظيفة اختبارات الاستعدادات.

فإذا استخدمنا اختباراً تحصيلياً فى فهم المقروء للتنبؤ بمدى إفادة الطالب من عملية تعليم اللغة العربية أو غيرها من المواد التى تتطلب ذلك الفهم، فإن هذا الاختبار يعد بمثابة اختبار تنبؤى.

ويمكن أيضًا أن نستخدم اختبارات التحصيل فى انتقاء الطلاب للالتحاق بالكليات، وكذلك انتقاء العاملين فى المجالات المختلفة. غير أن هذا الاستخدام ينبغى أن يبنى على أسس منطقية اعتماداً على ارتباط درجاتها بدرجات محك مستقبلى مثل النجاح فى الكلية، أو فى مهنة معينة. ولكى يكون التنبؤ باستخدام هذه الاختبارات جيداً ينبغى أن تقيس أنشطة مماثلة لمتطلبات دراسة أو مهنة معينة.

فدرجات اختبارات التحصيل لطلاب نهاية مرحلة التعليم الثانوى ربما نستطيع باستخدامها التنبؤ بدرجات الطلاب فى السنة الأولى الجامعية بدرجة أفضل من درجات اختبارات الذكاء. وبالطبع ربما تزداد القيمة التنبؤية إذا استخدمنا نوعى الاختبارات معًا.

كما نستطيع باستخدام بطاريات الاختبارات التحصيلية انتقاء أفضل المتقدمين لشغل وظائف معينة تتطلب معارف ومهارات يمكن قياسها بهذه الاختبارات. وكذلك يمكن استخدامها فى مجالات التوجيه والتدريب المهنى لاتخاذ قرارات تتعلق بالالتحاق بدراسة معينة أو برنامج تدريبى معين، ولكن هذا يتطلب التحقق من الصدق التنبؤى للاختبار التحصيلى المستخدم.

(٣) طريقة الاستجابة : يمكن أيضًا تصنيف الاختبارات التحصيلية تبعًا لطريقة الاستجابة المرجوة، فكثيراً من هذه الاختبارات تعتمد على الورقة والقلم، وتنقسم إلى اختبارات تعتمد على التعرف Recognition ، وأخرى تعتمد على الاسترجاع Recall. فاختبارات التعرف تشتمل عادة على مفردات اختيار من متعدد، أو صواب أو خطأ، أو مزاوجة حيث تتطلب الاختيار من بين بدائل معطاة، أو تحديد ما إذا كانت عبارة معينة صحيحة أو خطأ. ويستخدم هذا النوع من المفردات بكثرة فى اختبارات التحصيل المقننة، كما يستخدمه المعلم فى الاختبارات التى يُعدها بنفسه لطلابه. أما اختبارات الاسترجاع فتشتمل على أسئلة تتطلب تكملة أو إجابة قصيرة أو أسئلة مقال. وفى هذه الحالة يقدم السؤال ويكتب الطالب الإجابة التى يراها مناسبة.

كما أن بعض اختبارات التحصيل تقتصر على قياس معرفة الحقائق والمبادئ المتعلقة بمادة دراسية معينة دون الاهتمام الكافى بقياس قدرة الطالب على استخدام

وتوظيف هذه المعرفة فى مواقف تتطلب التفكير المثمر وحل المشكلات. وربما يرجع ذلك إلى صعوبة بناء اختبارات تشتمل على مفردات اختيار من متعدد لقياس هذه المستويات المعرفية العليا. غير أنه إذا علمنا أن المفردات الاختبارية التى تقيس هذه المستويات تسهم فى انتقال أثر التعلم المدرسى إلى مواقف أخرى متنوعة، فإن ذلك يشجع على الاهتمام بقياس تحققها لدى الطلاب.

كما أن اختبارات التحصيل يمكن أيضًا أن تقيس المهارات العملية. فالاختبارات العملية أو اختبارات الأداء تقيس مهارة الطالب فى تطبيق ما تعلمه بأن يقوم بتشغيل جهاز، أو يجرى تجربة، أو يعزف على آلة موسيقية، أو يؤدى تمارين رياضية معينة، وهكذا.

(٤) طريقة تصميم وبناء الاختبار: سبق أن أوضحنا فى الفصل الأول أن أدوات القياس يمكن تصنيفها إلى أدوات يُعدها المعلم وأخرى مقننة، ونؤكد هنا أن هذا التصنيف ينطبق على اختبارات التحصيل. فالاختبارات التحصيلية الصفية يقوم المعلم أو مجموعة من المعلمين فى المدرسة بإعدادها لطلابهم لقياس الأهداف التعليمية للمواد الدراسية، وبذلك يكون محتواها مرتبطًا ارتباطًا مباشرًا بما تم تدريسه فى المدرسة.

أما اختبارات التحصيل المقننة فيقوم بإعدادها خبراء القياس ومؤسسات نشر الاختبارات، وتُستخدم فى مدارس مختلفة نظراً لأنها تقيس أهدافا تعليمية عريضة، ومحتواها يكون متسعًا. فهى تقيس معارف اتفق خبراء التربية على ضرورتها لمعظم الطلاب فى مادة دراسية لفرقة معينة، ويتم تجريب الاختبار على عينة عشوائية كبيرة نسبيًا من الطلاب بحيث تكون خصائص هذه العينة محددة تحديدًا دقيقًا، وتُحسب معايير الاختبار لاستخدامها فى مقارنة أداء الطالب بأقرانه.

(٥) تفسير درجات الاختبار : أوضحنا فى الفصل السادس أن درجات الاختبار يمكن تفسيرها فى ضوء معيار نسبى أو محك أداء مطلق. فاختبارات التحصيل التى تهتم بمقارنة أداء الطالب بمعيار جماعة مرجعية لتحديد مركزه النسبى أو ترتيبه بين أقرانه تعد من نوع الاختبارات مرجعية الجماعة أو المعيار Norm - Referenced Tests أما اختبارات التحصيل التى تهتم بتحديد مدى تمكن الطلاب من مهارة محددة فى إطار محتوى دراسى معين، فإن درجته فى الاختبار يتم مقارنتها بمحك أو مستوى أداء محدد مسبقًا دون مقارنة درجته بمتوسط أداء جماعة مرجعية معينة ، وتعد هذه الاختبارات من

نوع الاختبارات مرجعية المحك Criterion - Referenced Tests ، ويتميز هذا النوع من الاختبارات بأن له قيمة تشخيصية لتحديد جوانب القوة والضعف في تحصيل الطالب؛ لذلك فإنها تكون أقرب إلى الاختبارات التي يُعدها المعلم لطلابه.

وسوف نناقش فيما يلي كيفية بناء كل من هذين النوعين من الاختبارات نظراً لأهميتهما في قياس وتقويم التحصيل الدراسي أو الأكاديمي، والتحقق من فاعلية البرامج التعليمية والتدريبية.

أولاً: اختبارات التحصيل مرجعية المحك:

يُستخدم هذا النوع من الاختبارات في كثير من الأغراض التربوية التي تتعلق بتقويم كفايات الأفراد وما يمتلكون من معارف ومهارات وظيفية مكتسبة من برامج تعليمية أو تدريبية محددة الأهداف والنواتج.

وتتميز هذه الاختبارات بأنها تصف السلوك وصفاً دقيقاً بما يُمكن من تشخيص جوانب القوة والضعف واقتراح أساليب العلاج. لذلك انتشر استخدام الاختبارات مرجعية المحك في المؤسسات التربوية والتدريبية في الولايات المتحدة الأمريكية وغيرها من دول العالم.

فهناك اهتمام متزايد في هذه الدول بتطوير برامج التعليم والتقويم الذي يستند إلى الكفايات Competencies والموديولات التعليمية Instructional Modules ، وبرامج التعليم بالاستعانة بالحاسوب Computer - Assisted Instruction حيث يحصل الطلاب على قدر كبير من التعليم من خلال التفاعل المتبادل مع الحاسوب. فكل طالب يتعلم وفقاً لسرعته الذاتية وإمكاناته في برنامج تعليم فردي. وعقب انتهائه من تعلم وحدة صغيرة أو مهارة معينة يجري البرنامج اختباراً مرجعي المحك لتحديد مدى تقدم الطالب. فإذا حقق المستوى المطلوب يمكن أن ينتقل إلى الوحدة الصغيرة أو المهارة التالية، أما إذا لم يحقق هذا المستوى يمكن أن يعيد تعلم الوحدة مرة أخرى حتى يصل إلى المستوى المطلوب مما يجعله يُوظف أقصى إمكاناته في التعلم.

كما أن المعلم يمكنه الإفادة من طرق بناء هذا النوع من الاختبارات في تطوير الاختبارات الصفية التي يكتبها لطلابه مما يزيد من فاعليتها في تقويم كفاياتهم ومكتسباتهم وتشخيص الصعوبات التي يواجهها غالبيتهم. ويفيد تطبيق هذه الاختبارات

بصورة دورية فى متابعة التقدم الدراسى للطلاب وحفزهم على التعلم المستمر، كما يفيد فى تقويم فاعلية برنامج تعليمى معين استناداً إلى البيانات المتجمعة من التطبيق الدورى لهذه الاختبارات مما يقدم مؤشراً لنمو الطلاب ومدى تحقيقهم للأهداف التعليمية المتعلقة بهذا المحتوى الدراسى.

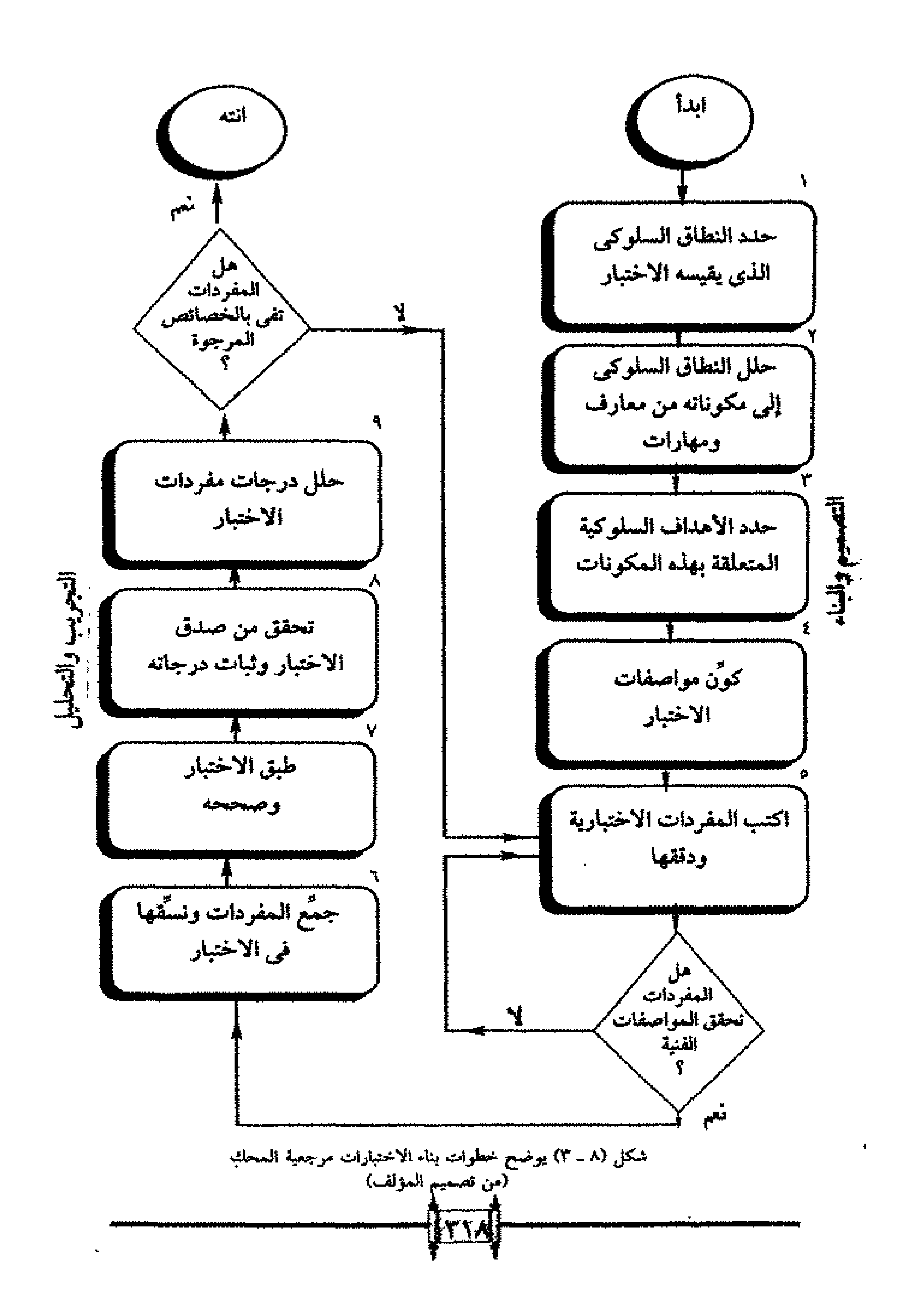
ونظراً لأهمية هذا النوع من الاختبارات فقد قام مؤلف هذا الكتاب منذ عام ١٩٨٠ ببناء مجموعات من هذه الاختبارات فى مجالات متعددة، كما أعد مرجعًا متخصصًا للاختبارات التشخيصية مرجعية المحك (علام ، ١٩٩٥) وأدلة معلم لبناء هذه الاختبارات (علام ، ١٩٨٣، ١٩٩٥، ١٩٩٧)، كما قام تلاميذه بإعداد عدد آخر من هذه الاختبارات فى مواد دراسية مختلفة فى رسائلهم للماجستير والدكتوراه.

لذلك سنناقش بإيجاز كيفية بناء هذه الاختبارات وتفسير نتائجها، وعلى القارئ المهتم الرجوع إلى المراجع السابقة وغيرها من المراجع المتخصصة للاستزادة فى هذا المجال.

وكذلك يمكن الرجوع إلى الكتاب المصدر الذى أعده مركز دراسات التقويم التابع لجامعة كاليفورنيا (CSE, 1989) المتعلق بالاختبارات مرجعية المحك الذى يشير إلى مجموعة متنوعة من هذه الاختبارات فى المجالات المختلفة التى تم نشرها، كما يشير إلى تعليق المركز على كل اختبار منها.

خطوات بناء الاختبارات مرجعية المحك :

يتطلب بناء الاختبارات مرجعية المحك إعداد خطة تفصيلية دقيقة قبل البدء الفعلى فى كتابة المفردات أو الأسئلة التى يشتمل عليها الاختبار. ويوضح الشكل التخطيطى (٨ ـ ٣) التالى خطوات بناء هذه الاختبارات.



شكل (٨ ـ ٣) يوضح خطوات بناء الاختبارات مرجعية المحك
(من تصميم المؤلف)

ويتضح من شكل (٨ ـ ٣) أن بناء الاختبارات مرجعية المحك يمر بخطوات متعددة سوف نناقشها فيما يلى:

الخطوة الأولى : تحديد النطاق السلوكى الذى يقيسه الاختبار وتحليله إلى مكوناته من معارف ومهارات :

فالتعريف الدقيق للنطاق السلوكي Behavioral Domain الذى يقيسه الاختبار مرجعي المحك يعد متطلبًا أساسيًا من متطلبات هذا النوع من الاختبارات، فالنطاق السلوكي يعد بمثابة الإطار المرجعي الذى تنسب إليه درجة الفرد فى الاختبار. ونقصد بالنطاق السلوكي المعارف والمهارات المحددة تحديدًا دقيقًا بما يمكننا من معرفة ما يستطيع الفرد أداؤه وما لا يستطيع أداؤه. ويذكر بابام Popham أن التحديد الدقيق للنطاق السلوكي يعد الركيزة الأساس لمفهوم المحك Criterion.

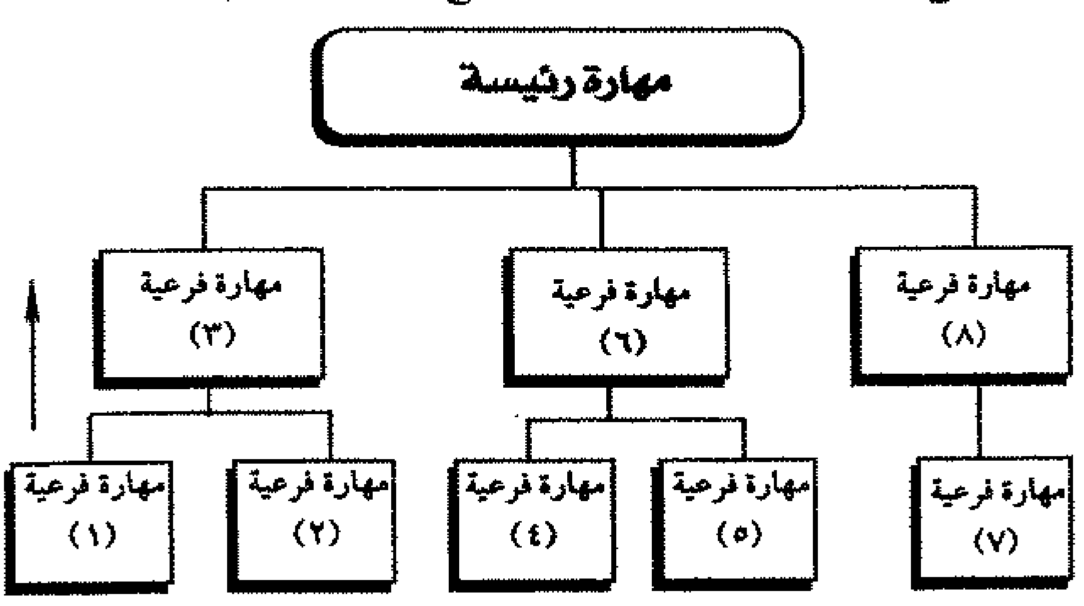
وبالطبع تختلف النطاقات السلوكية اختلافًا ملحوظًا فى درجة تعقدها ونوعية الأداء الذى تتضمنه ومتطلبات النمو أو التعلم.

وكلما زاد تعقد النطاق السلوكي تطلّب تحليله إلى مكونات يمكن قياسها جهدًا ملحوظًا. فنطاق مهارة فهم المقروء ربما يكون أكثر تعقدًا من نطاق مهارة رسم الخرائط، وهذا بدوره ربما يكون أصعب من نطاق معرفة جمع الأعداد البسيطة. ويفضل فى مثل هذه الحالات الاستعانة بخبراء المادة الدراسية أو اختصاصيى تصميم البرامج التدريبية فى تحديد النطاق السلوكي المراد قياسه، وتحليله إلى مكوناته وما تشتمل عليه من معارف ومهارات وسلوكيات.

وهناك طرق متعددة لإجراء هذا التحليل لعل أهمها طريقة تحليل بنية المحتوى Structural Analysis التى اقترحها سكاندورا (Scandura, 1977) وتحليل المهام Task Analyis الذى اقترحه جاجنى وبريجز (Gagne & Briggs, 1974).

ويهدف هذا التحليل لتحديد المتطلبات أو المكونات السلوكية Prerequisite Behavior اللازم توافرها لدى المتعلم لكى يحقق الكفاية أو المهارة الرئيسة التى يقيسها الاختبار، ونكرر عملية تحليل هذه المتطلبات حتى نصل إلى السلوك المدخلى Entery Behavior للمتعلم الذى اكتسبه بالفعل. ثم تُرتب هذه المتطلبات أى المعارف والمهارات المساعدة ترتيبًا هرميًا بنائيًا بحسب إسهاماتها فى تكوين المهارة الرئيسة. ويُمثل السلوك المدخلى قاعدة البناء الهرمى ثم تتدرج مستويات السلوك من الأبسط إلى الأكثر صعوبة حتى نصل إلى قمة البناء الذى يمثل الكفاية أو المهارة الرئيسة.

ويوضح شكل (٨ ـ ٤) التالى تخطيطًا لنواتج مثل هذا التحليل:



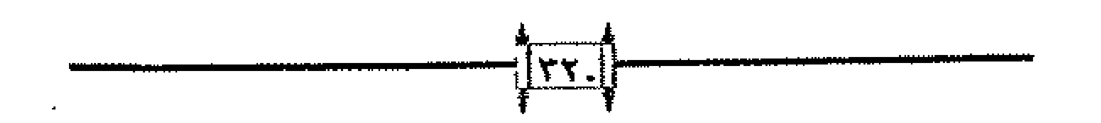
شكل (٨ ـ ٤) يوضح التخطيط الهرمى البنائى لمهارة رئيسة افتراضية

وإجراء هذا التحليل يتطلب معرفة أنماط التعلم الرئيسة ومستوياتها المتدرجة ونواتجها لكى يكون التحليل منطقيًا ومتكاملاً (علام، ١٩٩٥).

الخطوة الثانية: صياغة الأهداف السلوكية المتعلقة بنواتج تحليل النطاق السلوكى:

فالمهارات الفرعية التى نتجت عن تحليل النطاق السلوكى ينبغى صياغتها صياغة سلوكية بحيث يمكن قياسها بمفردات اختبارية. وينبغى أن تشتمل عبارات الأهداف السلوكية على وصف السلوك المتوقع بأفعال تشير إلى نشاط معين يمكن ملاحظته Action Verbs ، مثل يصنف، يميز، يقترح، يستخرج، ينسق، يخطط، يبرر، يدقق، يتعرف، وهكذا. ويراعى أن يتناسب الفعل مع نمط التعلم المطلوب ومستواه المعرفى.

كما يجب أن تشير هذه العبارات إلى المحتوى المرجعى الذى سوف يتعامل معه الفرد أثناء اختباره، مثل محتوى الصيغ الجبرية، أو المعارك الحربية، أو برامج الحاسوب، أو الصيغ الكيميائية، أو خدمات الإنترنت، أو دوائر التلفاز، وما شابه



ذلك. ويمكن أيضًا أن تُحدد الشروط والإجراءات التي يتم تقييم الأداء في ضوئها، أي تحديد ما إذا كان تحقق الهدف يتطلب سرعة أو دقة معينة، أو استجابة تحريرية أو شفهية أو حركية، وغير ذلك من الشروط التي تتطلبها طبيعة الهدف السلوكي.

الخطوة الثالثة: تكوين المواصفات التفصيلية للاختبار:

يختلف الاختبار مرجعي المحك عن غيره من الاختبارات في تحديد مواصفاته تحديدًا تفصيليًا دقيقًا.

فنوع مفردات الاختبار وطبيعة المثيرات التي تشتمل عليها وما تتضمنه من محتوى، وخصائص الاستجابات وتعليماتها، وطريقة ومحكات تقدير درجاتها ينبغي أن تحدد تحديدًا دقيقًا. والهدف من هذه التفاصيل المطولة تحديد النطاق السلوكي للاختبار بحيث يمكن الاسترشاد به في بناء مفردات متجانسة تقيس جميعها هدفًا سلوكيًا محددًا وارتباطها به ارتباطًا مباشرًا بما يجعل هناك تزاوج بين الهدف السلوكي والمفردة الاختبارية التي تقيسه، ويضمن صدق محتواها وصدق محتوى الاختبار ككل. وربما يتطلب ذلك أيضًا التحقق من صدق محتوى المواصفات التفصيلية للاختبار بواسطة مجموعة من خبراء المجال الدراسي أو التدريبي المعين.

ويرجع الاهتمام بهذه المواصفات التفصيلية إلى أنه يُستند إليها في كتابة مفردات متجانسة وبناء اختبارات متوازية أو متكافئة على الرغم من اختلاف الاختصاصيين الذين يقومون بإعداد هذه الاختبارات.

الخطوة الرابعة: كتابة أو تكوين مفردات الاختبار:

فمفردات الاختبار مرجعي المحك تنبثق مباشرة من مواصفات الاختبار، فإذا كانت هذه المواصفات مبرمجة فإنه يمكن استخدام الحاسوب في تكوين مفردات الاختبار Computer Items Generation ،وهناك بعض الأساليب التي يمكن الاسترشاد بها في إجراء هذه البرمجة مثل صيغ المفردات Item Forms ، وقواعد التحويلات السيمانتية Semantic Transformations ، والجمل الراسمة Mapping Sentences ، وغيرها من الأساليب التي ناقشها المؤلف في مراجع أخرى (علام ١٩٨٦ ، ١٩٩٥) أما إذا استرشد بمواصفات تفصيلية دون محاولة برمجتها ، فإنه ينبغي كتابة المفردات Item Writing بواسطة القائمين ببناء الاختبار.

وعندئذ ينبغى تصميم خطة معاينات المفردات لانتقاء نوع وعدد المفردات التى تُمثل النطاق السلوكى تمثيلاً كافيًا مع مراعاة محتوى النطاق ومكونات المهارة التى يتضمنها. فإذا كان النطاق السلوكى يقيس مهارة فهم المقروء مثلا، فإنه ينبغى مراعاة طول القطع المقروءة وعددها، وصعوبة كلماتها، ومجال الموضوعات التى تتناولها، والمثيرات أو المؤشرات التى تتضمنها، وغير ذلك من المتغيرات. ويجب الاسترشاد بهذه الخطة عند كتابة مفردات الاختبار نظراً لأن الإجراءات المنظمة لمعاينة النطاق السلوكى Domain Sampling تُعد من أهم خصائص الاختبارات مرجعية المحك.

فتمثيل مفردات الاختبار للنطاق السلوكى الشامل فى ضوء هذه الإجراءات يجعل من الممكن الاستدلال على درجة الفرد فى هذا النطاق من درجته فى عينة المفردات التى يشتمل عليها الاختبار.

ويتطلب كتابة مفردات الاختبار مهارة كبيرة وتمكُّن من محتوى النطاق السلوكى الذى يقيسه الاختبار.

وهناك نوعان من المفردات تستخدم فى قياس الجانب المعرفى والمهارى، أحدهما يتطلب إجابة مقيدة والآخر يتطلب إجابة حرة. ومن أمثلة النوع الأول مفردات الاختيار من متعدد، والصواب أو الخطأ، والمزاوجة. ومن أمثلة النوع الثانى أسئلة المقال، والإجابة القصيرة، والتكملة. ويتطلب كتابة كل من هذه الأنواع مراعاة بعض الشروط الفنية وبخاصة المفردات التى تتطلب إجابة مقيدة لقياس المهارات المعرفية العليا. وينبغى تدقيق المفردات عقب الانتهاء من كتابتها بواسطة مجموعة من خبراء المجال الذى يقيسه الاختبار وخبراء القياس والتقويم.

وقد أوضح المؤلف شروط وفنيات صياغة الأنواع المختلفة من المفردات بأمثلة من مختلف المجالات الدراسية فى أكثر من مرجع آخر (علام ١٩٨٣، ١٩٩٥، ١٩٩٦، ١٩٩٧).

الخطوة الخامسة: تجميع مفردات الاختبار:

بعد الانتهاء من كتابة المفردات وتدقيقها ينبغى تنسيقها فى ضوء الأهداف التى تقيسها لتكوين الاختبار. وهذا يتطلب مراعاة التسلسل المنطقى للمفردات وتجميع المفردات التى من نوع واحد معًا فى ضوء مكونات المهارة التى تم تحليلها ومستوياتها المعرفية. وينبغى أن يسبق كل نوع منها تعليمات واضحة ودقيقة تخبر الفرد كيف

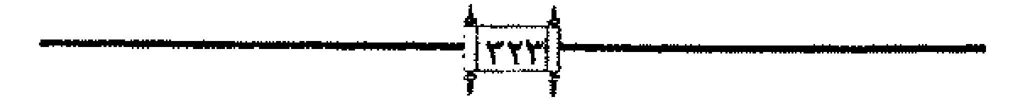
يستجيب وأين يدون استجابته، حيث إن هذه التعليمات تعد جزءًا من الاختبار. كما ينبغي تحديد زمن الاختبار إذا كان اختبارًا موقوتًا في ضوء طبيعة مفرداته ونوعها ومحتواها، ومن ثم يمكن طباعة الاختبار ومشتملاته.

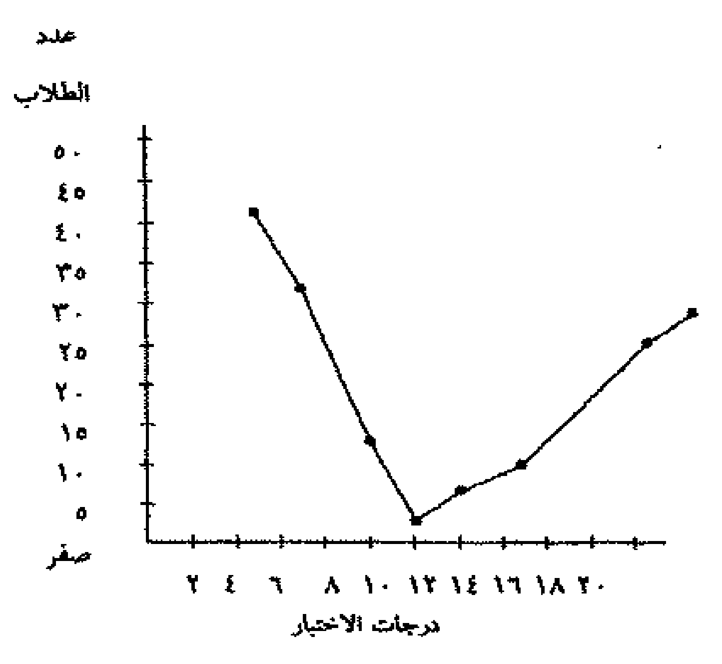
الخطوة السادسة: تجريب الاختبار ميدانيًا وتحليل مفرداته:

بعد الانتهاء من تكوين أو كتابة مفردات الاختبار وتجميعها ينبغي تجريب الاختبار ميدانيًا على عينة مناسبة أو أكثر من الأفراد لتعرُّف مدى ملاءمة المفردات للمختبرين ودرجة تمييزها بين الأفراد الذين حققوا هدفًا معينًا والأفراد الذين لم يتمكنوا من تحقيقه، وذلك لتقييم صدق المفردات في قياس النطاق السلوكي المتعلق بالأهداف المحددة. ويستطلب تحديد خصائص المفردات استخدام بعض الأساليب الإحصائية للبيانات المستمدة من التجريب الميداني للاختبار. غير أن هذه الأساليب ربما تختلف عن تلك التي ناقشناها في الفصل السابع المتعلقة بتحليل مفردات الاختبارات مرجعية الجماعة أو المعيار. ويرجع هذا الاختلاف إلى خصوصيات الاختبارات مرجعية المحك واستخداماتها، فهذا النوع من الاختبارات لا يهدف لإبراز الفروق الفردية في السمة المراد قياسها، وإنما يهتم بقياس تمكن الأفراد من نطاق سلوكي معين يتعلق بمهارة أو مهمة محددة نتيجة عملية التعليم أو التدريب. وهنا يفترض أن يحقق جميع الأفراد الأهداف التي يتضمنها النطاق السلوكي إذا كانت عملية التعليم أو التدريب فاعلة مما يجعل درجاتهم في الاختبار متقاربة. ونظرًا لأن أساليب تحليل مفردات الاختبارات مرجعية الجماعة أو المعيار ـ كما سبق أن ذكرنا ـ تعتمد على معاملات الارتباط التي تعتمد بدورها على تباين الدرجات، فإن هذه الأساليب لا تناسب استخدامات الاختبارات مرجعية المحك.

فإيجاد كل من معامل صعوبة المفردات ومعامل تمييزها في هذه الحالة ينبغي أن يكون مناسبًا للقرارات التي تتخذ في ضوئها. وهذه القرارات تتعلق بتصنيف الأفراد بحسب مستويات تمكنهم من النطاق السلوكي الذي تقيسه مفردات الاختبار. فإذا طبق الاختبار على مجموعة من الأفراد قبل بدء برنامج تعليمي أو تدريبي معين نتوقع انخفاض درجاتهم، بينما إذا طبق عليهم الاختبار نفسه أو صيغة مكافئة له عقب انتهاء البرنامج، فإننا نتوقع حصول غالبيتهم على درجات مرتفعة.

ويوضح شكل (٨ ـ ٥) التالي التوزيع المتوقع للدرجات قبل عملية التعليم وعقب الانتهاء منها:





شكل (٨ ـ ٥) يوضح توزيع افتراضي متوقع لدرجات ٥٠ طالبًا قبل عملية التعليم وعقب الانتهاء منها

ويتضح من شكل (٨ ـ ٥) أن توزيع الدرجات يتخذ شكلاً مفتوحًا إلى أعلى، حيث يشير الخط البيانى الأيسر إلى توزيع درجات الطلاب قبل عملية التعليم نظرًا لأنها درجات منخفضة، بينما يشير الخط البيانى الأيمن إلى توزيع درجاتهم عقب الانتهاء من عملية التعليم نظراً لارتفاع الدرجات.

ونقطة التقاء الخطين (الدرجة ١٢) تعد نقطة فاصلة للطلاب قبل عملية التعليم وبعدها. وتُمثل هذه النقطة درجة القطع Cut - off Score للاختبار، فالدرجات الأقل منهـا تدل على عدم التمكن من النطاق السلوكى الذى يقيسه الاختبار، والدرجات الأعلى منها تدل على التمكن.

ويعتمد تحليل مفردات الاختبارات مرجعية المحك على تقييم هذا التغير الذى حدث نتيجة عملية التعليم.

لذلك ينبغى إيجاد معامل صعوبة المفردات لمجموعة الأفراد قبل التعليم وعقب الانتهاء منه، أو للمجموعة التى تلقت التعليم والمجموعة التى لم تتلق التعليم. فما يهمنا فى هذه الحالة هو درجة حساسية المفردة Item Sensitivity لعملية التعليم أو لفاعلية المعالجة التعليمية. إذ نتوقع أن يكون هناك فرق موجب بين قيمة معامل صعوبة كل مفردة بعد التعليم وقبله.

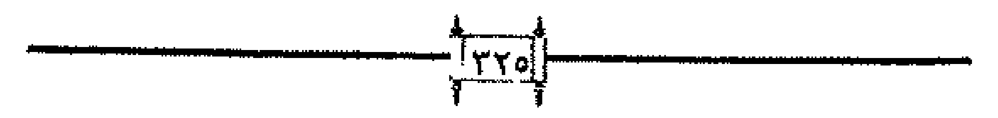
ولإيجاد معامل تمييز المفردات اقترح كوكس Cox وفارجاس Vargas أسلوبًا بسيطًا يعتمد على طرح نسبة عدد من أجاب إجابة صحيحة عن كل مفردة من مفردات الاختبار قبل بدء عملية التعليم من نسبة عدد من أجاب إجابة صحيحة عن كل من هذه المفردات عقب الانتهاء من عملية التعليم. وتتراوح قيم هذا الفرق بين - ١ ، + ١ . فكلما اقترب الفرق من (+ ١) دل ذلك على أن المفردة تُميز تمييزًا مرتفعًا بين الأفراد في مرتى تطبيق الاختبار أى قبل بدء عملية التعليم وعقب الانتهاء منها. وإذا اقترب الفرق من الصفر أو كان سالبًا دل ذلك على عدم فاعلية المفردة فى التمييز بين الأفراد مما يتطلب مراجعتها للتعرف على جوانب القصور فيها، والاسترشاد بآراء المختبرين حول فاعلية البرنامج التعليمى أو التدريبى بالنسبة لهم قبل إجراء التعديلات المناسبة على المفردة.

ومن الجدير بالذكر أن هناك العديد من الطرق الأخرى التى تستخدم فى تحليل مفردات الاختبارات مرجعية المحك يمكن أن يرجع إليها القارئ فى المصادر التى سبق أن أشرنا إليها.

الخطوة السابعة: التحقق من صدق الاختبار وثبات درجاته:

تختلف أساليب التحقق من صدق الاختبارات مرجعية المحك وثبات درجاتها اختلافًا ملحوظًا عن أساليب تقدير صدق وثبات الاختبارات مرجعية الجماعة أو المعيار التى ناقشناها بإسهاب فى الفصلين الرابع والخامس، وذلك على الرغم من عدم اختلاف المعنى العام لمفهومى الصدق والثبات فى كل من نوعى الاختبارات. فمفهوم الصدق يتعلق بمنطقية تفسير الدرجات وما يترتب عليها من قرارات صائبة، أما مفهوم الثبات فيتعلق باتساق الدرجات. وكما أن هناك أنواعًا مختلفة من الصدق والثبات فى الاختبارات مرجعية الجماعة أو المعيار، توجد أيضًا أنواع من صدق وثبات درجات الاختبارات مرجعية المحك. غير أننا سوف نقتصر هنا على الصدق المتعلق بالنطاق السلوكى Domain - Referenced Validity ، والثبات المتعلق باتساق تصنيف الأفراد بحسب درجة تمكنهم من نطاق سلوكى معين.

فالاختبار مرجعى المحك يقيس الأداء أو السلوك المتعلق بنطاق سلوكى من المعارف أو المهارات يُعد بمثابة السلوك المحك Criterion Behavior ، وكما سبق أن ذكرنا ينبغى تحديد السلوك المحك تحديدًا دقيقًا. ونقصد بصدق النطاق السلوكى للاختبار أن المحتوى والمهام التى تشتمل عليها عينة مفرداته تمثل السلوك المحك تمثيلاً كافيًا، فالنطاق السلوكى يشير عندئذ إلى هذا السلوك المحك.



ولعل هذا النوع من الصدق يشبه إلى حد ما صدق محتوى الاختبارات مرجعية الجماعة أو المعيار. غير أن مفهوم صدق المحتوى أكثر اتساعًا من مفهوم صدق النطاق السلوكى نظرًا لأن الاختبار مرجعى المحك يقيس عادة نطاقًا سلوكيًا محدودًا، مثل رسم أشكال هندسية، أو تكوين جمل صحيحة، أو إجراء عمليات حسابية، أو استنباط نتائج من تجربة عملية، وهكذا.

والتحقق من صدق النطاق السلوكي يتطلب تقييم إجراءات تحليل النطاق السلوكى إلى مكوناته من معارف ومهارات، ونواتج هذا التحليل (الخطوة الأولى فى بناء الاختبار التى سبق ذكرها)، وخطة معاينة مفردات الاختبار لهذه المكونات فى ضوء المواصفات التفصيلية للاختبار (الخطوتان الثالثة والرابعة) . ويعتمد هذا التقييم على أحكام الخبراء فيما يتعلق بنتائج هذه الخطوات للتحقق مما إذا كانت مواصفات النطاق السلوكى محددة تحديدًا تفصيليًا واضحًا ودقيقًا ومكتملاً ومرتبطًا ارتباطًا وثيقًا بمكونات هذا النطاق، وكذلك ما إذا كانت مفردات الاختبار تعد بمثابة عينة كافية وممثلة لهذه المكونات. لذلك فإن صدق النطاق السلوكى يتطلب وصفًا دقيقًا للنطاق المعين أو السلوك المحك، ومؤشرات عن مدى كفاية تمثيل مفردات الاختبار لهذا النطاق.

وللتحقق من اتساق تصنيف الأفراد المختبرين إلى متمكنين أو غير متمكنين، فإنه ينبغى تطبيق الاختبار عليهم وإعادة تطبيقه أو تطبيق اختبار آخر مكافئ له. ويقدّر الثبات بمقارنة نسبة عدد الأفراد الذين يتم تصنيفهم على أنهم متمكنين فى كل من مرتى التطبيق. وكلما زاد اتساق قرار التصنيف فى المرتين دل ذلك على ارتفاع درجة الثبات.

ولكن تصنيف الأفراد ينبغى أن يستند إلى درجة فاصلة أو درجة قطع محددة Cut - off Score كما ناقشناها فى الخطوة السادسة. كذلك ينبغى مراعاة أخطاء التصنيف الموجبة والسالبة التى أشرنا إليها فى الفصل الخامس فى سياق مناقشتنا للصدق التنبؤى باستخدام جداول التوقع. والأسلوب الإحصائى الذى يأخذ هذه الأخطاء بعين الاعتبار هو معامل كابا Kappa Coefficient، حيث يحدد نسب اتساق التصنيف وفقًا لدرجة قطع معينة، وتصحيح هذه النسب من أخطاء التصنيف التى تعزى إلى الصدفة. إذ ربما يكون قد تم تصنيف بعض الأفراد إلى متمكنين أو غير متمكنين عن طريق الصدفة.

والصيغة الرياضية لمعامل كابا كالتالى:

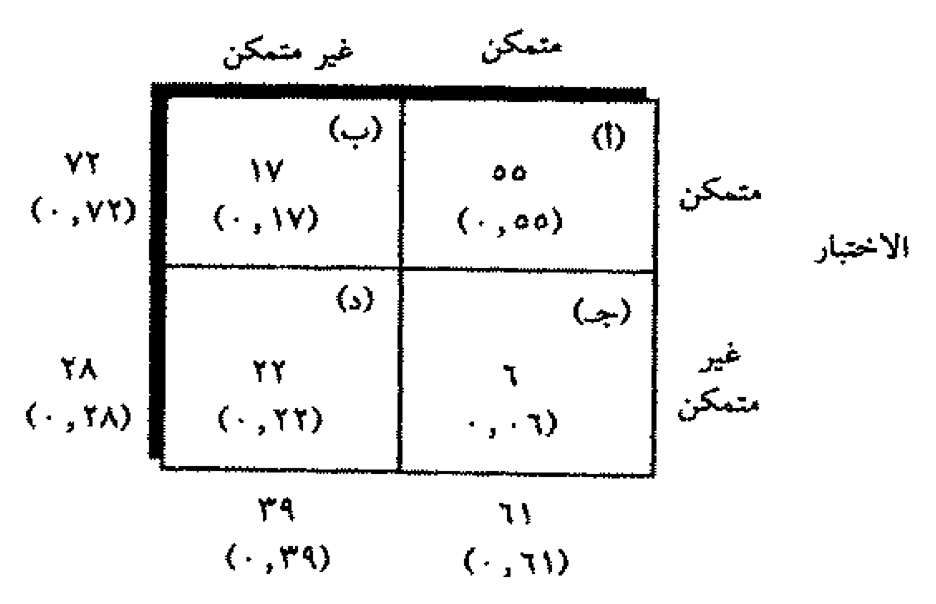


$$\text{معامل كابا } (K) = \frac{ل_{.} - ل_{م}}{١ - ل_{م}} \quad .... (٨ - ١)$$

حيث (ل.) ترمز إلى نسبة اتفاق التصنيف. وهى تساوى مجموع نسب أعداد الأفراد الذين يتم تصنيفهم فى المجموعة نفسها فى مرتى التطبيق ، (لم) ترمز إلى نسبة الاتفاق المتوقع عن طريق الصدفة، وهى تساوى مجموع النسب الهامشية لأعداد الأفراد الذين يتم تصنيفهم فى المجموعة نفسها فى مرتى التطبيق.

ولتوضيح ذلك نفترض أننا طبقنا اختباراً مرجعى المحك على مجموعة تتكون من ١٠٠ فرد، وأعدنا تطبيقه فى اليوم التالى، واخترنا درجة قطع ٩٠٪ ، وأجرينا تصنيفا لهؤلاء الأفراد فى مجموعتين إحداهما متمكنة والأخرى غير متمكنة استناداً إلى درجة القطع المحددة. وأوجدنا نسب أعداد من أجابوا إجابة صحيحة عن مفردات الاختبار فى مرتى التطبيق، ونسب أعداد من أجابوا إجابة خطأ.

ويوضح جدول (٨ ـ ١) التالى هذه النتائج:



إعادة تطبيق الاختبار

الاختبار	متمكن	غير متمكن	
متمكن	(أ) ٥٥ (٠,٥٥)	(ب) ١٧ (٠,١٧)	٧٢ (٠,٧٢)
غير متمكن	(جـ) ٦ ٠,٠٦)	(د) ٢٢ (٠,٢٢)	٢٨ (٠,٢٨)
	٦١ (٠,٦١)	٣٩ (٠,٣٩)	

جدول (٨ ـ ١) يوضح كيفية إيجاد قيمة معامل كابا (k)

فالخطوة الأولى: نوجد مجاميع النسب الهامشية لكل من الصفين، أى مجموع النسبتين فى الخليتين (أ) ، (ب) وكذلك (جـ) ، (د) أى ٠,٧٢ ، ٠,٢٨ ، وكذلك مجاميع النسب الهامشية لكل من العموديين أى مجموع النسبتين فى كل من الخليتين (أ)، (جـ) ، والخليتين (ب) ، (د) أى ٠,٦١ ، ٠,٣٩ ويلاحظ أن ٠,٥٥ من الأفراد تم تصنيفهم إلي متمكنين ، ٠,٢٢ تم تصنيفهم إلى غير متمكنين فى مرتى التطبيق.

وبذلك تكون ل. = ٠,٥٥ + ٠,٢٢ = ٠,٧٧

كما يلاحظ أن ٠,٧٢ ،٠,٦١ هما المجموعين الهامشيين للتصنيف «متمكن»، ٠,٢٨ ، ٠,٣٩ هما المجموعين الهامشيين للتصنيف «غير متمكن». وبذلك تكون

$$ل_{م} = (٠,٧٢)(٠,٦١) + (٠,٢٨)(٠,٣٩)$$

$$= ٠,٥٤٨٤$$

والخطوة الثانية: نعوض فى الصيغة (٨ ـ ١) لإيجاد قيمة معامل كابا (K) :

$$\text{معامل كابا (K)} = \frac{٠,٧٧ - ٠,٥٤٨٤}{١ - ٠,٥٤٨٤} = ٠,٤٩$$

ولتفسير هذه القيمة ينبغى ملاحظة أن قيم معامل كابا لا تزيد عن (+١) ولا تقل عن (-١)، ونحصل على القيمة الأولى عندما تكون النسب الهامشية متساوية فى مرتى التطبيق، أما القيمة الثانية فنحصل عليها إذا اختلف تصنيف الفرد فى المرتين.

ونظراً لأن القيمة التى حصلنا عليها فى هذا المثال ٠,٤٩ فإن هذا يعنى أن نسبة اتساق التصنيف منخفضة مما يدل على انخفاض قيمة الثبات.

وتتأثر قيمة معامل كابا باختلاف درجة القطع التى نختارها، وكذلك تتأثر بعدد مفردات الاختبار ومستوى صعوبتها، وعدد الأفراد المختبرين. لذلك ينبغى تدوين المعلومات المتعلقة بهذه المتغيرات إلى جانب قيمة معامل كابا للاسترشاد بها عند تفسير هذه القيمة. وينبغى الإشارة إلى أننا سبق أن ناقشنا مستويات الأداء ودرجات القطع فى نهاية الفصل السادس.

ويمكن الرجوع إلى المصادر التى أشرنا إليها فى هذا الفصل للاطلاع على مزيد من طرق تحديد هذه المستويات أو الدرجات.

استخدامات الاختبارات مرجعية المحك :

نشأت حركة القياس مرجعي المحك كما أوضحنا في مستهل هذ الفصل مرتبطة بفلسفة تربوية للتعليم تختلف عن فلسفة القياس مرجعي الجماعة أو المعيار، فقد تبين قصور هذا النوع الأخير من القياس في التقويم التشخيصي، وكذلك قصوره في تقويم فاعلية البرامج التعليمية من حيث مدى تحقيقها للأهداف المرجوة، وبذلك لا يساعد في اتخاذ قرارات تفيد في تطوير هذه البرامج وزيادة فاعليتها.

فالأساليب والممارسات التربوية التقليدية أدت إلى جعل وظيفة المؤسسات التعليمية تقتصر بدرجة كبيرة على تصنيف المتعلمين والحكم على ضعاف التحصيل بالفشل واستبعادهم تدريجيًا خلال المراحل الدراسية، وتشجيع القليلين على متابعة التعليم دون الاهتمام بتقديم العون الكافي لجميع الأفراد للتعلم الفاعل واستكمال التعليم. فالنظام التعليمي السائد يوحّد أساليب التدريس والمدة الزمنية للتعلم، وتُصمم الاختبارات بطريقة تجعل الطلاب الضعاف أو من يعانون من صعوبات التعلم يحصلون على درجات منخفضة حيث إن معيار الأداء في هذه الاختبارات يعتمد على الموازنة بالطالب المتوسط. وهذا يتنافى مع وظيفة المؤسسات التعليمية التي ينبغي أن يكون هدفها الجوهري هو التعليم والتعلم الذي هو حق للجميع. ومع التطورات الكبيرة الحادثة في المجالات العلمية والتكنولوجية يقل الطلب على العمالة الماهرة ويلاقي أمثال هؤلاء الطلاب الذين لم يتح لهم قدرًا كافيًا من التعليم صعوبات جمّة في الحصول على أعمال، حيث إن الأعمال المتاحة التي تكون مناسبة لهم محدودة جدًا. لذلك ينبغي أن ترتبط عمليات التعليم والتقويم بالأهداف التربوية المستقبلية المتسعة النطاق وتعمل على تحقيقها وتنمية إمكانات كل متعلم إلى أقصى حد ممكن بما يجعله قادرًا على مواجهة احتياجات ومتطلبات هذا القرن. ولعل حركة الاختبارات مرجعية المحك وما ارتبط بها من تقنيات وتطبيقات والتي أصبحت شائعة الاستخدام في العديد من الدول المتطورة وبعض الدول النامية وما ترتب عليها من منهجية جديدة تؤكد التكامل بين عملية التعليم وعملية التقويم من خلال الأهداف التربوية المستقبلية تعد من التطورات المهمة التي وجّه علماء القياس وخبراء التصميمات التعليمية جهودهم البحثية نحوها لتوسيع نطاق تطبيقاتها داخل الصف المدرسي.

تقييم الكفايات Competence Assessment:

إن منهجية القياس مرجعي المحك باعتبارها منهجية إديومترية مستحدثة أثرت تأثيرًا واضحًا فى النظر إلى التحصيل من منظور جديد، فأصبح ينظر إلى قياس التحصيل على أنه تقييم للكفايات Competence Assessment، أى تقويم القدرة على تنفيذ مهام محددة بنجاح بدلاً من اختزان بعض المعلومات والمعارف المتفرقة، وهذا يجعلنا نفسر التحصيل بأنه أداءً Performance أو قدرة على إنجاز شيء ما. فالكفاية تتضمن توسيع نطاق المعرفة والفهم وتنمية مهارات متعددة ومتكاملة يصعب قياسها بواسطة الاختبارات التى تصمم وفق الأساليب السيكومترية الكلاسيكية. وقد أسهم هذا التوجه في تطوير تصميمات ونماذج تعليمية مبنية على أساس الكفايات Competency-Based Models من أجل التمكن والإتقان في ضوء فلسفة تربوية مؤداها أن كل طالب مهما كانت قدرته يستطيع أن يتعلم تعلمًا فاعلا ويتقن الأهداف والكفايات المحددة إذا أتيح له الوقت الكافى والأسلوب التعليمى الذى يناسب مستوى قدرته (Carroll, 1970 Guskey, 1985)، وهذا يتطلب تنوعًا في أساليب التعليم وأدوات التقويم وتقنياتها. وقد استخدمت هذه التصميمات والنماذج التعليمية فى كثير من المجالات التربوية والمهنية في الدول المتطورة بهدف إكساب الكفايات وتقويمها تقويمًا بنائيًا وتشخيصيًا وختاميًا وتحسين الممارسات في هذه المجالات.

وتُعد حركة تقويم المهارات الأساسية Basic Skills Assessment من التطورات المرتبطة بالقياس مرجعى المحك، وتهتم هذه الحركة بالكفايات الوظيفية المتعلقة بالقراءة والكتابة والرياضيات والحاسوب، وكذلك بمهارات العمل والوظائف، وتقويم التحصيل المرتبط بالحياة الفعلية والمهام التى يواجهها المتعلم في تعامله اليومى تقويمًا مرجعى المحك.

كذلك من أهم تطبيقات الاختبارات مرجعية المحك منح التراخيص المهنية. ففى الولايات المتحدة الأمريكية حددت كثير من المنظمات المهنية اختبارات مدخلية ينبغى أن يجتازها الفرد المتقدم للعمل فى مهنة معينة أو يحقق حدًا أدنى من مستوى محدد للسماح له بمزاولة المهنة. كما يمكن أن يعاد تطبيق الاختبار على الفرد إذا أخفق فى المرة الأولى. وعادة تقيس هذه الاختبارات الكفايات المتعلقة بالأدوار المهنية، وتفسر درجات الفرد فى ضوء مستويات الأداء المحددة.

كما أن اختلاف أنواع المؤسسات التعليمية وأهدافها يتطلب الحصول على أدلة متعددة عن التحصيل وفق محكات محددة. ويمكن أن يتمثل ذلك في الموهبة الفنية، والإنتاج الأدبي، والقيادة، والتفكير العلمي، والمهارات البدنية، وخدمة المجتمع، وفنيات صنع الأشياء، ومهارة تحليل المشكلات المعقدة، مما يمكن قياسه وتقييمه باستخدام الاختبارات مرجعية المحك.

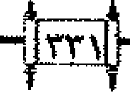
وكذلك يمكن الإفادة من هذه الاختبارات في تشخيص أخطاء مهارات العمليات Process Skills Errors المتعلقة بإجراءات تنفيذ مهمة معينة، وأخطاء فهم المقروء Reading Errors ، وأخطاء التحويل Transformation Errors التي تعوق المتعلم عن انتقاء العملية أو سلسلة العمليات اللازمة لحل المشكلة، وأخطاء فك الترميز Encoding Errors التي تجعله لا يستطيع التعبير عما توصل إليه من حل للمشكلة بصورة مكتوبة مقبولة على الرغم من معرفته للحل.

عقود الأداء Performance Contracting

تسعى كثير من الدول إلى ضبط جودة الأداء التربوي، وتحديد مستويات هذا الأداء بانتظام، وتعرُّف جوانب الضعف فيه وتشخيصه ومعالجته، والإفادة من ذلك في تطوير نوعية المناهج المدرسية والممارسات التربوية. لذلك ينادي كثير من المسئولين في المجال التربوي في بعض دول العالم المتطور وكذلك أولياء أمور الطلاب بتقويم نواتج العملية التعليمية في ضوء مدى تحقيقها للأهداف التربوية والمجتمعية. ويصبح المعلم مسئولا عن تقديم أدلة تبرهن على أن تلاميذه قد حققوا هذه الأهداف بعد مرور مدة دراسية محددة. وتُستمد هذه الأدلة من نتائج تطبيق اختبارات مرجعية المحك أثناء عملية التعليم وعقب الانتهاء منها. فهذه الاختبارات تتوافر فيها خصائص معينة تجعلها أكثر ملاءمة لقياس الأهداف التعليمية قياسًا مباشرًا في ضوء مستوى أداء محدد مسبقًا، كما أوضحنا في هذا الفصل.

ويرى مؤيدو حركة المسئولية التربوية Accountability في الولايات المتحدة الأمريكية وغيرها من الدول، أن عقود الأداء هي الوسيلة التي تضمن التزام جميع الأطراف المعنية بتحقيق الأهداف المرجوة للبرنامج التعليمي على أساس زمني محدد.

غير أنه لا توجد مؤشرات واضحة تدل على مدى فاعلية تطبيق هذا النظام في المجالات التعليمية المختلفة، وبخاصة أنه يتطلب تنفيذ أنظمة إدارية متطورة



تعتمد على التقنيات الحديثة، مثل تقويم ومراجعة البرامج Program Evaluation and Revision System.

مشاركة الجمهور في العمليات الإختبارية :

أدت التوجهات السابقة إلى زيادة مشاركة الجمهور في العمليات الاختبارية، وعمليات التقويم في بعض الدول المتطورة، وما ترتب على ذلك من حوار ومناقشات، ونقد وتشريعات، ومراجعة التنظيمات في ظل هذا المناخ الديمقراطي الذى أدى إلى حركة المسئولية التربوية Accountability Movement التى أشرنا إليها.

وعلى الرغم من الآثار الإيجابية لهذا التطور الذى أفادت هذه الدول من نتائجه في السنوات الماضية مثل زيادة عدالة الاختبارات وغيرها من أساليب تقويم التحصيل، وتوسيع نطاق المشاركات في تحديد محتوى الاختبارات مرجعية المحك، والأهداف التعليمية التى تقيسها هذه الاختبارات ، وعدم إساءة استخدامها، وتقديم معلومات أفضل للمختبرين ومستخدمى الاختبارات في ضوء هذه الأهداف إلا أن هذا لم يخل من مشكلات تتعلق باتخاذ قرارات اعتبارية أو اعتماد مخططات ربما تتنافى مع ممارسات التقويم السليم.

تقويم البرامج التربوية Educational Program Evaluation

يهدف تقويم البرنامج التربوى لتعرُّف مدى تحقيقه لأهدافه المرجوة. والاختبارات مرجعية المحك تفيد ليس فقط في التقويم الختامى للبرنامج لتعرُّف مدى فاعليته في تحقيق تلك الأهداف، وإنما تفيد بدرجة أكبر في التقويم البنائى للبرنامج وذلك بالكشف عن جوانب القوة والضعف في أهدافه وتصميمه وإجراءاته وأساليب تنفيذه بحيث يمكن تعديل مسار البرنامج بناء على نتائج هذه الاختبارات.

كما يمكن إعادة النظر في أهداف البرنامج أو محتواه إذا تبين عدم ملاءمتها لمستوى قدرات الطلاب والإمكانيات المادية والفنية المتوافرة.

ثانيا : اختبارات التحصيل مرجعية الجماعة أو المعيار :

يرتبط هذا النوع من الاختبارات ارتباطًا وثيقًا باختبارات التحصيل المقننة Standaradized Achievement Tests . وتُعد هذه الاختبارات أدوات قياس تربوية مهمة نظرًا لأنه يمكن أن تتخذ قرارات متعددة تتعلق بالأفراد والبرامج التربوية والتدريبية استنادًا إلى نتائجها.

وتختلف اختبارات التحصيل المقننة عن الاختبارات مرجعية المحك أو الاختبارات التي يكتبها المعلم لطلابه في أنها تعتمد اعتماداً أساسياً على إجراءات تقنين منظمة وجداول معايير يُستند إليها في مقارنة أداء فرد أو مجموعة من الأفراد بأداء جماعة مرجعية Norm Group مواصفاتها محددة بدقة.

لذلك تقوم بإعداد هذه الاختبارات مؤسسات متخصصة تحت توجيه وإشراف خبراء القياس التربوي واختصاصيي المناهج، وذلك لأنها تتطلب كثيراً من الإمكانات المادية والفنية. وينتشر استخدام الاختبارات التحصيلية المقننة في الولايات المتحدة الأمريكية بخاصة حيث يُنفق على بنائها واستخدامها ملايين الدولارات، وتقوم المؤسسات المتخصصة بنشر وتوزيع هذه الاختبارات في كافة أنحاء العالم. غير أننا نعاني في الوطن العربي من قلة هذه الاختبارات وندرة استخدامها في تقويم المستوى العام للطلاب والمدارس، وفي فحص وانتقاء العاملين بالمهن والوظائف المختلفة. وربما يرجع ذلك إلى عدم وجود مؤسسات عربية متخصصة يوكل إليها مهام إعداد هذه الاختبارات بأنواعها المختلفة التي سبق أن أشرنا إليها.

ففي الولايات المتحدة الأمريكية تقوم هذه المؤسسات بإعداد مئات من الاختبارات في المواد الدراسية المختلفة لقياس التحصيل العام لطلاب جميع المراحل التعليمية تكون عادة بطاريات من الاختبارات، وكذلك إعداد اختبارات مقننة لمهارات أساسية في المرحلتين الابتدائية والمتوسطة، مثل القراءة، وفهم المقروء، والهجاء، والعمليات الحسابية والهندسية، وتوظيف الأرقام والأشكال البيانية، وحل المسائل، وغيرها من مهارات التعبير والتواصل، والتعامل مع المقادير الكمية. كما تقوم بإعداد اختبارات مقننة في مجالات مهنية معينة تتطلب تقييماً لمعارف ومهارات العاملين فيها أو المتقدمين للالتحاق بها.

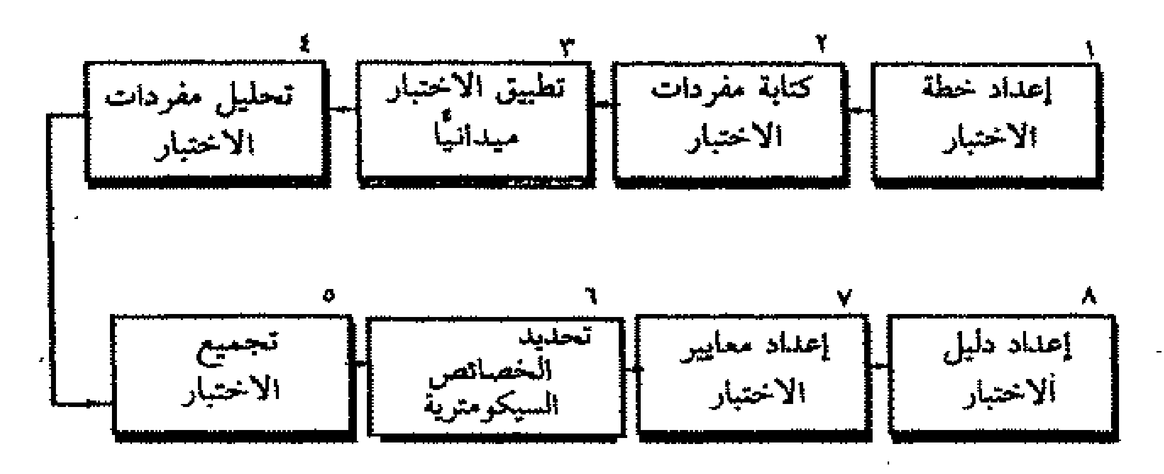
وتنتقي المدارس أو المؤسسات المهنية الاختبارات والمقاييس المقننة التي تناسب المرحلة الدراسية أو مجال العمل المطلوب، وخصائص الأفراد،والغرض من التقويم. ولا يكون للمعلم أو المسئول عن العمل دور في إعداد هذه الاختبارات، بل ولا يقوم بتصحيحها في كثير من الأحيان، وإنما تقوم المدرسة أو الجهة المعنية بإرسال أوراق الإجابة إلى المؤسسة الناشرة للاختبار حيث تُجرى عملية تصحيح الإجابات باستخدام أجهزة إلكترونية مخصصة لهذا الغرض، وذلك لأن هذه الاختبارات تشتمل عادة على مفردات اختيار من متعدد. وتُسجِّل هذه الأجهزة الدرجات الخام للأفراد في الاختبار، وكذلك المئينيات أو الدرجات المعيارية المعدلة أو التساعيات المعيارية المناظرة لهذه



الدرجات الخام لكي يتسنى مقارنة أداء كل فرد بأداء الجماعة المرجعية التي استُمدَّت منها معايير الاختبار. ويعاد إرسال النتائج إلى الجهة المعنية بأحد وسائل الاتصال العادية أو الإلكترونية.

خطوات بناء اختبارات التحصيل المقننة:

يمر بناء اختبارات التحصيل المقننة بخطوات متعددة. وعلى الرغم من أن هذه الخطوات ربما تختلف بحسب طبيعة الاختبار ونوعه واستخداماته، إلا أن هناك تسلسلا منطقيًا معينًا ينبغي مراعاته يوضحه الشكل التخطيطي (٨ ـ ٦) التالي:



شكل (٨ ـ ٦) يوضح التسلسل المنطقي لمراحل إعداد الاختبار التحصيلي المقنن

(من تصميم المؤلف)

وسوف نوضح فيما يلي هذه الخطوات بمزيد من التفصيل:

إعداد خطة الاختبار: فبناء الاختبار التحصيلي المقنن يتطلب العناية بإعداد خطة دقيقة ومنظمة يُستند إليها في تنفيذ الخطوات المتتالية الموضحة بشكل (٨ ـ ٦)، ويقوم بإعداد هذه الخطة مجموعة متكاملة من خبراء القياس والتقويم واختصاصيي المواد الدارسية أو المجالات المهنية والتدريبية ومستشاري المناهج. ولإعداد هذه الخطة ينبغي تحديد الغرض من الاختبار وفيم تُستخدم نتائجه. فقد سبق أن ذكرنا أن هناك أنواعًا متعددة من هذه الاختبارات، وتحديد النوع الذي تتبين الحاجة إليه يعد أمرًا ضروريًا. غير أنه في جميع الأحوال ينبغي تحديد محتوى الاختبار وذلك بفحص عينات من الكتب المدرسية في المجال المعين حيث إنها تعد مصدرًا أساسيًا للمقرر أو المنهج المدرسي للفرقة المستهدفة في النظام التعليمي في مختلف أنحاء القطر. وكذلك

استشارة خبراء المادة الدراسية والمعلمين المتميزين إما فرادى أو جماعة. أما إذا كان المجال يتعلق بمهن معينة فإنه ينبغي تعرف مواصفات هذه المهن ومتطلباتها الأساسية استرشاداً بآراء اختصاصيي هذه المهن والمطبوعات المتعلقة بمواصفاتها. والهدف من ذلك تحديد الموضوعات والمشروعات والإجراءات التي تتضمنها، والوقت اللازم لتعلم كل منها. وهذا يتطلب تحليلاً دقيقاً للمحتوى ومتعلقاته، والاتفاق على نواتج هذا التحليل من جانب هؤلاء الخبراء، نظراً لأن الاختبارات التحصيلية المقننة تستخدم على نطاق واسع. ويتم في ضوء ذلك إعداد جدول مواصفات الاختبار Table of Specifications ، وهو جدول ثنائي البعد تُعيّن فيه المجالات أو الموضوعات المتعلقة بالمحتوى على البُعد الأفقي، والمعارف والمهارات المرجو تحقيقها في مستويات معرفية مختلفة حيث تصاغ في عبارات عامة على البُعد الرأسي مثل معرفة المصطلحات والحقائق، فهم المقروء ، حل المسائل، اقتراح تفسيرات، تقييم مهام أو مطالب، وهكذا. وتدون في خلايا الجدول أوزانا نسبية لكل من المحتوى والمهام وفقاً لأهميتها، والزمن المناسب لتعلمها، وإسهامها في الموضوعات اللاحقة، مع مراعاة تجنب التحيزات الفردية.

ويوضح جدول (٨ ـ ٢) التالي هيكل جدول مواصفات افتراضي لاختبار تحصيلي مقنن في القراءة:

مجالات المحتوى

المعارف والمهارات	كلمات	نحو	معلومات	تعبير وظيفي
معرفة				
فهم				
تطبيق				
تحليل				

جدول (٨ ـ٢) يوضح هيكل جدول مواصفات اختبار تحصيلي مقنن في القراءة

ويُعد جدول (٨ ـ ٢) بمثابة خطة تنظيمية لمجالات المحتوى ومستويات المعرفة التي يهدف الاختبار لقياسها باستخدام عينة ممثلة من المفردات التي يُقترح عددها في خلايا الجدول في ضوء الاعتبارات السابقة.

كتابة مفردات الاختبار: بعد الانتهاء من إعداد جدول المواصفات يبدأ فريق الخبراء في كتابة مفردات الاختبار التي تقيس المعارف والمهارات الموضحة على البعد

الرأسي في ضوء المحتوى الموضح على البُعد الأفقي. وتُعد هذه الخطوة من أصعب خطوات بناء الاختبارات المقننة نظراً لما تتطلبه من جهد كبير، وما تستغرقه من وقت طويل في كتابة مفردات ذات نوعية جيدة لقياس المعرفة أو المهارات المرجوة قياساً صادقًا.

ويتطلب ذلك تنظيم ورش عمل مكثفة لمجموعات الخبراء في أيام متتالية لكتابة المفردات وصياغتها ومناقشتها ومراجعتها مراجعة دقيقة وإعادة صياغتها وانتقاء أفضلها. كما يستطلب أحيانًا أن يكتب كل خبير على حدة مفردات مقترحة يجرى مراجعتها وتنقيحها في ورش العمل. وعادة يتم كتابة عدد من المفردات أكبر مما هو مدون في خلايا جدول المواصفات نظراً لأن بعض هذه المفردات ربما يستبعد عند فحصها وتدقيقها أو بعد تجريبها، كما أنه عادة تكون هناك صيغ متكافئة للاختبار مما يتطلب توافر عدداً كبيراً من المفردات.

وعلى الرغم من كثرة استخدام مفردات الاختيار من متعدد في الاختبارات المقننة، إلا أنه ليس هناك ما يمنع من استخدام أنواع أخرى من المفردات أو الأسئلة. فقد سبق أن ذكرنا أن المفردات الاختبارية يمكن تصنيفها إلى مفردات تتطلب إجابة مقيدة وتسمى مفردات موضوعية، ومفردات تتطلب إجابة حرة أو مفتوحة، ومفردات تتطلب أداءات عملية، ويعتمد انتقاء نوع المفردة على المهام أو المهارات المراد قياسها. غير أن مركز الخدمات الاختبارية بالولايات المتحدة الأمريكية Educational Testing Services (ETS) الذي يعد من المراكز المرموقة في بناء مختلف أنواع الاختبارات التي يتم نشرها في كل دول العالم أصبح يستخدم أنواعًا مختلفة من المفردات عند بناء الاختبارات المقننة. إذ يستخدم مفردات الاختيار من متعدد التي تتطلب من المستجيب وضع علامة معينة في ورقة الإجابة تشير إلى اختياره لأحد البدائل التي تشتمل عليها المفردة. فهذه المفردات تتميز بإمكانية تطويعها بحيث تقيس المستويات المعرفية العليا، ويمكن أن يجيب الفرد عن عدد كبير منها في وقت أقل مما يجيب فيه عن سؤالين أو ثلاثة أسئلة مقالية، كما أنها مُحددة الإجابة مما يجعل تصحيحها أيسر وأقل كلفة وأكثر اتساقًا، بل ويتم تصحيحها عادة باستخدام الحاسوب.

أما أسئلة المقال فإنها تناسب قياس الأصالة والطلاقة والفكر المنتج، وتوضيح استيعاب الحقائق والتعبير عن ذلك في إطار تنظيمي مناسب، وعلى الرغم من عمقها في القياس إلا أن ذلك يكون على حساب دقة القياس. فتقدير الدرجات يشوبه عدم الاتساق، مما يستطلب الحرص والعناية عند صياغة أسئلة المقال بحيث يكون الإطار المرجعي للإجابة عنها محدداً وفي الوقت نفسه يسمح للفرد بأن يُعبِّر عما في حوزته من معارف تتعلق بمضمون السؤال مما يجعل عملية تقدير الدرجات أكثر موضوعية.

ويمكن تصميم أنواع من المفردات تضم ميزات المفردات الموضوعية وأسئلة المقال كما في الاختبارات التي تقيس التعبير اللغوى، وتسمى المفردات المتداخلة خطيًا Interlinear Items ،حيث يُطلب من المستجيب أن يقوم بتحرير قطعة نثرية تحتوى على بعض الأخطاء سواء في النحو أو الصياغة أو الفكرة التى تتناولها القطعة.

كما يمكن تصميم أنواع أخرى من المفردات غير التقليدية لتقييم نوعية مهارات معينة، مثل مهارة العلاقات الاجتماعية لدى الأطفال. وهنا يُقدم للأطفال موقف يتطلب حل مشكلة بطريقة جماعية كأن يطلب منهم تصميم مجسم معين باستخدام عدد محدد من المكعبات التى توزع عليهم. ويسجل القائم بالملاحظة سلوك كل منهم أثناء العمل المشترك، ويترجم هذه الملاحظات إلى درجات يمكن في ضوئها تصنيف مجموعة الأطفال في قسمين فيما يتعلق بمهارة العلاقات الاجتماعية، مثل المجموعة المتعاونة في التخطيط، والمجموعة المتعاونة في التنفيذ.

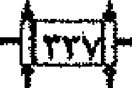
وكذلك يمكن تصميم مفردات من نوع آخر غير تقليدى تتعلق بالكفاءة الموقفية Situational Proficiency Items ،حيث يتركز الاهتمام على تقييم كفاءة الأفراد فى أداء مهام محددة تحاكى مواقف العمل الفعلية، وتسمى هذه المفردات «مفردات فى السلة In - Basket Items» . وتُستخدم هذه المفردات عادة لتقييم العاملين فى الوظائف التى تتطلب كفاءات ومهارات عليا معينة مثل الوظائف الإدارية أو القيادية.

فلتقييم المهارات الإدارية تصمم مفردات تتضمن محاكاة المواقف الإدارية بما فى ذلك المراسلات والمذكرات والأدلة الإرشادية والتقارير التى تتطلب اتخاذ قرارات إدارية متعددة. كما يُستخدم هذا النوع من المفردات الاختبارية فى دراسة أساليب تدريب وانتقاء المتقدمين للالتحاق بهذه الوظائف.

أى أنه يمكن تحوير وتطوير أنواع المفردات المألوفة وابتكار أساليب جديدة لقياس الأهداف التعليمية والتدريبية التى تتناولها الاختبارات التحصيلية المقننة.

تطبيق الاختبار ميدانيا وتحليل مفرداته : بعد إعداد خطة الاختبار وكتابة مفرداته، فإن الخطوة التالية تطبيق الاختبار لتجريبه ميدانيًا وتحليل مفرداته. فمهما بلغت عناية الخبراء فى إعداد خطة الاختبار، ومهما كانت خبرتهم وحذقهم فى بناء أنواع المفردات الاختبارية، فإن هذا لا يضمن أن المفردات سيكون وقعها الميدانى مماثل لما توقعه الخبراء.

لذلك ينبغى إعداد صيغة أولية للاختبار Pre-Test تشتمل على عدد من المفردات أكبر مما هو مطلوب وذلك لتجريبها ميدانيًا، ويطبق الاختبار على عينة أو عينات من الأفراد تماثل المجموعة الفعلية التى سيُطبق عليها الاختبار فى صيغته النهائية فيما بعد.



لذلك يجب أن تمثل العينات المناطق الجغرافية، والمستويات الاجتماعية والاقتصادية والذكور والإناث، وغير ذلك من الخصائص الديموجرافية.

وتعد هذه الخطوة من الخطوات المهمة في بناء هذا النوع من الاختبارات لأنه سيجرى بناء على نتائج هذا التطبيق تحليل مفردات الاختبار باستخدام الأساليب الإحصائية التى سبق أن أوردناها تفصيلاً فى الفصل السابع. ويؤدى هذا التحليل إلى إلقاء الضوء على صعوبة وتمييز كل مفردة.

ويفيد ذلك في تحليل فاعلية مكونات المفردات من أجل تحديد جوانب القصور فيها وتعديلها وحذف المفردات غير الصالحة، وربما إضافة مفردات جديدة وتجريبها ميدانيًا مرة أخرى، كما أنه يُفيد فى الحكم على الأنواع الجديدة أو المبتكرة من المفردات مثل التى أشرنا إليها من حيث إمكانية الاستمرار فى استخدامها، وتعرُّف أفضل استخداماتها، وما إذا كانت تقيس شيئًا يصعب قياسه باستخدام الأنواع المألوفة من المفردات، ومدى فاعليتها فى ذلك، وعددها المناسب الذى ينبغى استخدامه فى الاختبار لجعل الدرجات أكثر ثباتًا. وبذلك يمكن انتقاء أفضل المفردات المتعلقة بالأهداف المراد قياسها لكى يضمَّها الاختبار فى صورته النهائية.

وينبغى مراعاة أنه يفضل أن يتراوح مدى صعوبة المفردات بين السهولة والصعوبة لكى تميِّز المفردات تمييزًا أفضل بين المختبرين.

كما ينبغى أثناء التجريب الميدانى للمفردات تحديد الإجراءات المقننة لتطبيق الاختبار لكى تضمَّها تعليمات الاختبار عندما يوضع فى صورته النهائية.

تجميع الاختبار: بعد إجراء تحليل المفردات والتحقق من فاعليتها ينبغى تجميع الاختبار فى صورته النهائية. وهذا يتطلب التحقق من أن المفردات المنتقاة تمثل جميع خلايا جدول المواصفات فى ضوء الأوزان النسبية لكل خلية منها. كما ينبغى إعادة مراجعة الاختبار مرة أخرى بواسطة أحد اختصاصيي المادة التى يقيسها الاختبار للتحقق من عدم تحيز مفردات الاختبار لمناطق معينة، مثل المدن والريف أو لنوع الفرد وغير ذلك من عوامل التحيز.

كما ينبغى التحقق من أن عدد المفردات التى يشتمل عليها الاختبار مناسب للزمن الذى يتم تحديده عن طريق تطبيق الاختبار على مجموعة مماثلة من الأفراد مع مراعاة ما إذا كان الاختبار من نوع اختبارات القوة Power Test أو اختبارات السرعة Speed Test. فإذا كان الاختبار من النوع الذى يعتمد على القوة ينبغى أن يسمح لنسبة تبلغ حوالى ٩٠٪ من الأفراد أن يحاولوا الإجابة عن جميع المفردات، أما إذا كان من النوع الذى يتطلب السرعة فإنه يتوقع أن تحاول نسبة أقل الانتهاء من إجابة جميع

مفردات الاختبار. وفى أى الأحوال ينبغى حساب متوسط الزمن اللازم للإجابة عن الاختبار.

كما ينبغى إجراء عملية تكافؤ للصيغ المختلفة من الاختبار بحيث لا تختلف درجات الفرد من صيغة إلى أخرى اختلافًا دالا.

ويعد هذا التوازن فى عدد المفردات وصعوبتها وتمييزها وملاءمتها للمختبرين من العمليات الدقيقة التى يجب مراعاتها.

تحديد الخصائص السيكومترية للاختبار: وتتضمن هذه الخطوة تقدير ثبات درجات الاختبار، أى اتساق درجاته إذا أعيد تطبيق الاختبار على مجموعة الأفراد نفسها بعد مرور مدة زمنية مناسبة، وكذلك تقديم أدلة تتعلق بتكافؤ صيغتى الاختبار. أى إيجاد قيمة كل من معامل الاتساق ومعامل التكافؤ، وكذلك التحقق من الاتساق الداخلى لمفردات الاختبار إذا تبينت الحاجة إلى أدلة عن هذا النوع من الثبات.

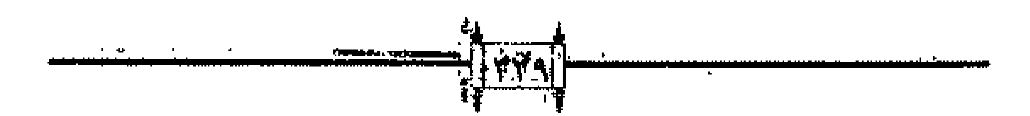
كذلك ينبغى التحقق من صدق نتائج الاختبار فى تحقيق الغرض الذى بُنى الاختبار من أجله. فبعض اختبارات التحصيل المقننة تستخدم فى أغراض الانتقاء والتسكين والتشخيص والتنبؤ، فهذه الاستخدامات تتطلب التحقق من الصدق المرتبط بمحك الذى يعتمد على بيانات إمبيريقية.

ولكن كثيراً من هذه الاختبارات تستخدم فى تقويم إنجازات الطلاب وفاعلية عملية التعليم، لذلك فإن محتوى الاختبار يعد العنصر الأساسى. وهذا يتطلب التحقق من صدق المحتوى لتقدير مدى تعلق مفردات الاختبار بالأهداف التعليمية العامة التى يهدف الاختبار لقياسها،ومدى تمثيلها لخلايا جدول مواصفات الاختبار. وقد سبق أن ناقشنا مفهومى الثبات والصدق بإسهاب فى الفصلين الرابع والخامس.

إعداد معايير الاختبار: بعد الانتهاء من تجريب الاختبار وتحليل مفرداته وانتقاء أفضل المفردات وتجميعها فى الاختبار، وتحديد خصائصه السيكومترية، فإن الخطوة التالية إعداد معايير الاختبار.

فعلى الرغم من الجهد الكبير الذى يبذل فى انتقاء المفردات وفى تكوين الاختبار، إلا أن الاختبار لا يستفاد منه ما لم يكن هناك إطار مرجعى يستند إليه فى تفسير الدرجات تفسيراً له معنى ودلالة.

لذلك لابد من إعداد معايير الاختبار التى تحقق ذلك. فهذه المعايير تمكّننا من مقارنة الدرجة أو الدرجات التى يحصل عليها الفرد بمتوسط درجات الجماعة المرجعية



أو عينة التقنين Norm Group ، كما تمكّننا من المقارنة بين درجات الأفراد فى المجالات المختلفة. لذلك فإن العناية بانتقاء عينة أفراد الجماعة المرجعية التى تستمد منها معايير الاختبار تعد من الأمور الأساسية فى بناء الاختبارات المقننة.

وقد سبق أن ناقشنا بالتفصيل خصائص الجماعة المرجعية، ومعايير الاختبارات وأنواعها وكيفية حسابها فى الفصل السادس. وأوضحنا فى هذا الفصل المفاهيم المتعلقة بالمعايير المحلية، والمعايير الوطنية،ومعايير مؤسسة معينة وفئات خاصة. وعلى الرغم من أن الاختبارات التحصيلية المقننة تُستخدم على مستوى وطنى مما يتطلب أن تكون عينة التقنين ممثلة لجميع المناطق الجغرافية وتراعى المتغيرات الديموجرافية الأخرى، إلا أنه يمكن أيضًا أن تحدَّد معايير محلية للاختبار أو معايير تتعلق بجماعات معينة من الأفراد.

وتتنوع معايير الاختبارات المقننة وفقًا لطبيعة الاختبار واستخداماته، فكثير من هذه الاختبارات تستخدم معظم أنواع المعايير التى ناقشناها فى الفصل السادس، حيث يجرى تحويل الدرجات الخام إلى درجات تائية، أو تساعيات معيارية،أو مئينيات،أو مكافئ العمر الزمنى، أو مكافئ الفرق الدراسية، وتُعرض نتائج هذا التحويل فى جداول معايير أو صفحات تربوية (بروفيل Profile)، أو أشكال بيانية وغير ذلك من الأساليب المبتكرة فى عرض معايير الاختبارات:

إعداد دليل الاختبار: إن بناء الاختبار وإعداد معاييره ليس نهاية المطاف، وإنما لابد من إعداد دليل أو أدلة الاختبار، وأوراق الإجابة، والملاحق المعينة لمستخدمى الاختبار بحيث يكون لديهم معلومات كافية يمكنهم فى ضوئها الحكم على ما إذا كان الاختبار يناسب الغرض الذى يسعون إليه، وما إذا كانت نتائجه يمكن استخدامها استخدامًا مناسبًا.

وعادة يلتزم ناشرو الاختبارات المقننة فى إعدادهم لأدلة هذه الاختبارات بالمعايير التى حددتها الجمعية الأمريكية لعلم النفس (APA) المتعلقة بالاختبارات التربوية والنفسية وأدلتها. وقد أوضحنا ذلك بالتفصيل فى الفصل الثانى.

فجميع الاختبارات المقننة ينبغى أن يصاحبها أدلة Test Manuals تزوِّد مستخدم الاختبار بمعلومات تفصيلية تتعلق بكيفية استخدام الاختبار، وتفسير درجاته، وخصائص مفرداته، وصدقه، وثباته، وتعليمات تطبيقه وتصحيحه. أى أن دليل الاختبار يعد المصدر الرئيسى لمستخدم الاختبار الذى يود أن يحصل على معلومات محددة عن اختبار معين. لذلك ينبغى أن يكون الدليل حديثًا ومكتملاً.

وبانتهاء جميع هذه الخطوات يصبح الاختبار صالحًا للنشر. غير أن عملية بناء الاختبارات المقننة لا تنتهى بمجرد نشر الاختبار، وإنما تتطلب المراجعة والتدقيق المستمر في ضوء ما يستجد من مطالب. وما تكشف عنه التطورات فى المجالات التربوية والنفسية، وما تتوصل إليه الدراسات والبحوث في مختلف هذه المجالات. فالاختبار الذى يبدو مُرضيًا اليوم ربما لا يكون كذلك غدًا، إذ ربما يلزم مراجعته أو بناء اختبار آخر بديلاً عنه مع تكرار جميع الخطوات السابقة التى تتضمن إعداد خطة الاختبار، وكتابة مفرداته، وتجريبه ميدانيًا،وتحليل مفرداته، وتجميعه، والتحقق من صدقه وثبات درجاته، وإعداد معاييره وأدلته التى يُسترشد بها فى تطبيقه وتصحيحه وتفسير درجاته.

وهذا يؤكد ما سبق أن ذكرناه فى مستهل هذا الجزء من ضرورة وجود مؤسسات متخصصة فى الوطن العربى يوكل إليها مهام بناء الاختبارات المقننة بأنواعها المختلفة، وكذلك الاختبارات مرجعية المحك ، واختبارات الكفايات فى مختلف المجالات والمهام والأعمال.

بعض أوجه النقد التى تُوجَّه للاختبارات التحصيلية المقننة:

على الرغم من مزايا الاختبارات التحصيلية المقننة واستخداماتها المتعددة فى مختلف المجالات التربوية والمهنية فيما يتعلق بإمكانية مقارنة أداء الأفراد فى مجال معين أو مجالات متنوعة بأداء أقرانهم، ومقارنة التحصيل والأداء بين الصفوف والمدارس المختلفة، وتتبع التقدم الدراسى للطلاب والنمو المهنى للعاملين، وتقييم فاعلية المناهج الدراسية، والإرشاد والتوجيه التربوى والمهنى، إلا أن علماء القياس وجّهوا كثيرًا من النقد لهذا النوع من الاختبارات لأسباب متعددة ذكرنا بعضًا منها فى الفصل الثانى تحت عنوان « الأخطاء الناجمة عن إساءة استخدام الاختبارات والمقاييس النفسية والتربوية » . وفيما يلى بعض أوجه النقد الأخرى:

(١) تشتمل بطاريات الاختبارات التحصيلية المقننة عادة على مفردات اختيار من متعدد فقط حيث يصعب استخدامها فى قياس قدرة الطالب على تنظيم وعرض الأفكار، ونقد الموضوعات، وتحليل وتركيب المعلومات وعرضها عرضًا متكاملاً، وغير ذلك من المهارات العقلية العليا التى ينبغى العناية بتنميتها لدى الطلاب. وبالطبع لا نستطيع إنكار عيوب الأسئلة التى تتطلب إجابات مفتوحة، مثل أسئلة المقال وغيرها من الأسئلة التى تعتمد على التقديرات الذاتية، والتى تقيس هذه المَهارات، حيث إنها لا تمثل المحتوى المراد قياسه تمثيلاً كافيًا بسبب الزمن الذى تتطلبه الإجابة عنها، وصعوبة

تصحيحها وعدم اتساق درجاتها ، مما يجعلها لا تناسب البرامج الاختبارية التي تُطبق على نطاق واسع وإن كانت تصلح للتطبيق في نطاق المدرسة. غير أن الاتجاه نحو تطوير هذا النوع من الأسئلة وابتكار أنواع جديدة من المفردات الاختبارية وتضمينها في اختبارات التحصيل المقننة ربما يعالج هذه المشكلة ولو معالجة جزئية.

(٢) تقيس هذه الاختبارات في كثير من الأحيان نواتج تعليمية تختلف عن النواتج التي يبغي المعلم تحقيقها في سلوك طلابه، فهي تقيس ما يهم خبراء الجهة أو المؤسسة المتخصصة الناشرة للاختبار بغض النظر عن الأهداف التربوية التي يهتم المعلم في منطقة أو مدرسة معينة بتحقيقها. وبالطبع يصعب على هذه المؤسسة بناء اختبار تحصيلي مقنن يناسب جميع المناطق وجميع المدارس ويمكن تطبيقه على جميع الطلاب في ظروف موحدة ومحددة.

(٣) يرتبط بالمشكلة السابقة مشكلة أخرى تتعلق بما إذا كانت المناهج المدرسية تتحدد بناء على الاختبارات التحصيلية المقننة، وما إذا كان المعلم يلجأ إلى التدريس من أجل الاختبار. فإذا كان تقويم المعلم يعتمد على نتائج طلابه في هذه الاختبارات، فإن المعلم سوف يؤكد في عملية التدريس ما يحتويه الاختبار ويغفل الموضوعات الأخرى التي ربما تكون أكثر أهمية، وذلك لكي يبدو تدريسه فاعلاً.

(٤) تُبنى اختبارات التحصيل المقننة على فلسفة تربوية تؤكد إبراز الفروق الفردية بين الطلاب، وجعل توزيع درجاتهم في الاختبار يتخذ شكل التوزيع الاعتدالي الذي تستند إليه معايير هذه الاختبارات. ولعل الاتجاه إلى استخدام الاختبارات مرجعية المحك أدى إلى تخفيف حدة هذه المشكلة، حيث إن هذه الاختبارات تؤكد أن الغرض من العملية التعليمية إكساب كل طالب الكفايات أو المعارف والمهارات،وتنمية قدراته واستعداداته،وتقويم هذه المكتسبات تقويمًا مباشرًا دون الحاجة إلى التركيز على الفروق الفردية بين الطلاب وبغض النظر عما اكتسبه كل منهم بالفعل. فإذا حقق كل طالب الأهداف المرجوة من العملية التعليمية فإننا نستطيع عندئذ الحكم على فاعليتها، وإبراز الفروق الفردية إذن في مثل هذه المواقف التعليمية لا يكون مرغوبًا ولا مطلوبًا.

لذلك نجد أن المؤسسات المتخصصة في بناء اختبارات التحصيل المقننة أصبحت تكرِّس جانبًا كبيرًا من جهودها وإمكاناتها في بناء اختبارات مرجعية المحك في

مختلف المجالات الدراسية والتدريبية والمهنية لقياس مدى تحقق الأهداف المتعلقة بها، وتشخيص جوانب الضعف التى ربما تعوق المتعلم عن اكتساب المعارف والمهارات المرجوة.

وأحيانا تُنتقى بعض مفردات اختبارات التحصيل المقننة وتُستخدم كمفردات مرجعية المحك. فالمفردات التى تقيس مهارة معينة يمكن استخدامها فى تقويم هذه المهارة بحيث إذا أجاب الطالب عنها إجابة صحيحة دل ذلك على اكتساب هذه المهارة دون الحاجة إلى تفسير درجته استناداً إلى معايير اختبار التحصيل المقنن الذى تُعده الجهة الناشرة، وإنما يمكن تحديد مستوى الأداء أو محك النجاح المطلوب الذى يُستند إليه فى تقويم الطالب. فمفردات اختبارات التحصيل المقننة تعد مفردات عالية الجودة نظراً لمرورها بفحص منظم ومراجعة دقيقة مما يجعلها صالحة لقياس الجوانب المعرفية.

وفحص إجابة كل طالب عن كل مفردة أو مجموعة من مفردات الاختبار مرجعى المحك على حدة بدلاً من الاهتمام بالدرجات الكلية فقط أو درجات الاختبارات الفرعية يقدم للمعلم معلومات تفصيلية عن كل طالب، ربما يصعب الحصول عليها باستخدام اختبار التحصيل المقنن. وبذلك يمكن الإفادة من كل من نوعى الاختبارات وإحداث تكامل فيما بينهما فى عملية تقويم الطلاب وإثراء عملية التعليم.

خلاصة

تهدف اختبارات التحصيل الدراسي أو الأكاديمي لقياس وتقويم مستوى الكفايات التي حققها المتعلم نتيجة للخبرات التعليمية، وعلى الرغم من أن هذه الاختبارات قد مرت بتغيرات متلاحقة، إلا أن استناد عملية التقويم إلى أساليب علمية وإجراءات منهجية أدى إلى تطوير أساليب وتقنيات بناء هذه الاختبارات.

فقد أصبح هناك أنواع متعددة من اختبارات التحصيل تُستخدم في اتخاذ قرارات تربوية تتعلق بالأفراد والبرامج. فاختبارات التحصيل المقننة بأنواعها المختلفة تُستخدم في وضع تقديرات للطلاب ومقارنة الأداء بين صفوف فرقة معينة في مدارس أو مناطق مختلفة، وتُقدم أدلة متجمعة عن نمو الطلاب تفيد في الإرشاد التربوي واتخاذ قرارات الانتقاء والتصنيف والتسكين وكذلك في تقويم فاعلية البرامج التعليمية. لذلك تقوم ببناء هذه الاختبارات مؤسسات متخصصة نظراً لكلفتها المرتفعة وما تتطلبه من خبرات فنية متميزة. وتعتمد عملية البناء على خطة تنظيمية دقيقة تحدد الخطوات المتتالية لإعداد مفردات الاختبار وتجريبها ميدانيًا وتحليلها،وكذلك أساليب التحقق من صدق وثبات درجات الاختبار، وإعداد معاييره التي يستند إليها في تفسير الدرجات.

أما الاختبارات مرجعية المحك فإنها تعد من التطورات المعاصرة التي تهدف لتقويم الكفايات المتعلقة بمختلف المجالات الدراسية والتدريبية والمهنية في ضوء مستوى أو محك أداء يتم تحديده مسبقًا دون الاهتمام بتحديد المركز النسبي للفرد بين أقرانه في المجال الذي يقيسه الاختبار.

لذلك يُعتنى في بناء هذه الاختبارات بتحديد النطاق السلوكي الذي يشتمل على المعارف والمهارات المرجوة تحديداً دقيقًا بحيث يمكن ملاحظة السلوك وقياسه. لذلك تختلف أساليب تحليل مفردات هذا النوع من الاختبارات وأساليب تقدير صدق الاختبار وثبات درجاته عن الأساليب المتبعة في الاختبارات مرجعية الجماعة أو المعيار. ويعتمد استخدام أي من النوعين على الغرض من استخدام الاختبار، وطبيعة القرار المراد اتخاذه نظراً لأن النوعين يختلفان في كيفية تفسير درجاتهما.

لذلك يمكن الإفادة من كل من النوعين بحيث يكون هناك تكامل فيما بينهما لإثراء عملية التقويم وترشيد القرارات التربوية استناداً إلى بيانات ومعلومات متنوعة.

غير أن هذا يتطلب تدريب المعلمين والمسئولين عن اتخاذ القرارات المتعلقة بالأفراد على كيفية انتقاء الاختبارات المناسبة وإجراءات تطبيقها وتقييم نتائجها للإفادة المثلى من اختبارات التحصيل بأنواعها المختلفة.

الفصل التاسع

قياس الذكاء الإنسانى

* مقدمة
* قياس الذكاء من منظور تاريخى
* مفهوم الذكاء والمشكلات المتعلقة بتعريفه
* الذكاء كمفهوم وصفى
* الذكاء كمجموعة من القدرات المترابطة
* الذكاء كعامل عام
* الذكاء كعوامل متعددة
* الذكاء كبنية ثلاثية الأبعاد
* الذكاء كنظام هرمى من العوامل
* الذكاء ومدخل تجهيز المعلومات ومعالجتها
* قياس الذكاء الإنسانى
* اختبارات الذكاء العام
* أولاً : اختبارات الذكاء الفردية
* مقياس ستانفورد ـ بينيه للذكاء
* مقاييس ويكسلر للذكاء
* مقارنة بين مقاييس ويكسلر ومقياس ستانفورد ـ بينيه
* بعض مقاييس الذكاء الفردية الأخرى
* ثانياً : اختبارات الذكاء الجماعية
* أنواع اختبارات الذكاء الجماعية
* أمثلة لبعض اختبارات الذكاء الجماعية
* اختبارات ذكاء جماعية لفظية متعددة المستويات
* اختبارات الذكاء الجماعية غير اللفظية
* تعقيب عام على مقاييس الذكاء
* بعض القضايا المتعلقة بقياس الذكاء
* بناء اختبارات الذكاء
* بعض محددات الذكاء والعوامل المؤثرة فيه
* ثبوت نسبة الذكاء
* خلاصة

مقدمة :

عرضنا في الفصل السابق طرق قياس التحصيل الدراسي والكفايات باستخدام الاختبارات مرجعية الجماعة أو المعيار، والاختبارات مرجعية المحك، واتضح لنا أن هذه الاختبارات تهدف لقياس المعارف والمهارات المتعلقة بمجال دراسي أو تدريبي محدد. غير أن اقتصار هذه الاختبارات على قياس الأداء الحالي للأفراد في مواقف مدرسية أو تدريبية معينة يجعلها محدودة الفائدة في التنبؤ بالأداء المستقبلي، وبخاصة إذا كانت المعارف والمهارات المرجو اكتسابها في المستقبل تختلف عما اكتسبه الفرد في الوقت الحالي، أو إذا كان صدق اختبارات التحصيل المستخدمة وثبات درجاتها منخفضًا. فالأداء المستقبلي يصعب التنبؤ به باستخدام متغير منبئ واحد، وهو درجات اختبار التحصيل، وإنما يتطلب استخدام متغيرات متعددة لكي يتسنى اتخاذ قرارات صائبة تتعلق بالأفراد استنادًا إلى الدرجات المتنبأ بها. فالذكاء أو القدرة العقلية العامة، والاستعدادات الخاصة المتعلقة بمجالات معينة تعد من المتغيرات المعرفية الأساسية التي يمكن استخدامها إلى جانب متغير التحصيل الدراسي في التنبؤ بالأداء المستقبلي للأفراد.

ونظرًا لأن متغير الذكاء أكثر عمومية من متغير الاستعدادات الخاصة فسوف نفرد هذا الفصل لتوضيح مفهوم الذكاء وما يتعلق بتعريفه من مشكلات منهجية وتطبيقية، وطبيعة الذكاء ونماذجه النظرية، وأساليب وأدوات قياسه وطرق بنائها، ومحددات الذكاء، والعوامل المؤثرة فيه. وتتناول مفهوم الاستعدادات الخاصة في الفصل العاشر. ولكي يتضح مفهوم الذكاء بدرجة أفضل سوف نستهل الفصل الحالي بتوضيح تطور هذا المفهوم وأساليب قياسه من منظور تاريخي في إطار تطور حركة القياس العقلي.

قياس الذكاء من منظور تاريخي :

على الرغم من أن ابتكار اختبارات الذكاء ربما يُعدُّ أحد التطورات السيكولوجية المهمة في القرن العشرين، إلا أن محاولات التفكير في طبيعة الذكاء لها تاريخ قديم.

فالذكاء حتى بداية القرن التاسع عشر كان يعد من المفاهيم الفلسفية التي لا تخضع للدراسة العلمية. ومن أوائل المحاولات التي أدت إلى وضع المعالم الرئيسة لقياس الذكاء قياسًا كميًّا على أسس علمية منظمة جهود كل من عالم البيولوجي الإنجليزي فرانسيس جالتون Galton (١٨٢٢ ـ ١٩١١) وهو من أتباع داروين Darwin البارزين، وعالم النفس الأمريكي جيمس ماكين كاتل Cattell (١٨٦٠ ـ ١٩٤٤).

وقد درس كاتل علم النفس التجريبي في جامعة ليبتزج في ألمانيا، وكان واحدا من تلاميذ ويلهلم فندت Wundt مؤسس معمل علم النفس التجريبي في ألمانيا عام ١٨٧٩، وأجرى بحوثه بعد حصوله على درجة الدكتوراه بالاشتراك مع جالتون، ثم انتقل إلى جامعة كولومبيا في الولايات المتحدة الأمريكية رئيسًا لقسم علم النفس، وكان أول من استخدم مفهوم «الاختبار العقلي Mental Test» عام ١٨٩٠. غير أنه اعتبر القدرة العقلية العامة تنعكس في التمييز الحسي وزمن رد الفعل وغير ذلك من الوظائف النفسية البسيطة.

وقد أفاد كاتل من جهود جالتون واهتمامه بالأساليب الإحصائية في استخدام هذه الأساليب كوسيلة مهمة للبحث في مجال الفروق الفردية.

وفي عام ١٩٠٤ طلب وزير التعليم في فرنسا من عالم النفس ألفرد بينيه Binet وزميله الطبيب النفسي سيمون Simon اقتراح كيفية التعرف على الأطفال الذين لا يستطيعون التعلم بكفاءة في الفصول المدرسية العادية. لذلك قام كل من بينيه وسيمون ببناء أول مقياس للذكاء يشتمل على ٣٠ مشكلة مرتبة ترتيبًا تصاعديًا ـ بحسب درجة صعوبتها ـ وتتطلب القدرة على الحكم، والفهم، والتعليل. ونشر هذا المقياس عام ١٩٠٥، وتمت مراجعته وتعديله عامي ١٩٠٨ ، ١٩١١ استنادًا إلى البيانات الإمبيريقية التي استمدت من تطبيق المقياس على طلاب المدارس الفرنسية. وأصبح المقياس يشتمل على عدد أكبر من الاختبارات الفرعية التي تم تصنيفها في مستويات عمرية مختلفة.

وبذلك استند المقياس في قياسه للذكاء إلى الوظائف العقلية العليا بدلاً من الوظائف الحسية البسيطة التي بينت دراسات كاتل عدم ارتباطها بالتحصيل الدراسي. واقترح بينيه وسيمون مفهوم «العمر العقلي Mental Age » الذي يحدد أداء الفرد بالنسبة لأقرانه في نفس مجموعته العمرية استنادًا إلى درجات المقياس. فإذا كان أداء طفل مساويًا متوسط أداء أطفال الثامنة من العمر مثلاً، فإن عمره العقلي يكون مساويًا ٨ سنوات، وقد يختلف عمر الطفل العقلي عن عمره الزمني سواء بالزيادة أو النقصان. ولعل مفهوم العمر العقلي يُعد من الإسهامات المهمة في مقياس الذكاء بعد تعديله.

غير أن ويلهلم شترن Stern وهو أحد علماء النفس الألمان يبدو أنه أول من اقترح مفهوم نسبة الذكاء (IQ) Intelligent Quotient التي تدل على النسبة بين العمر العقلي إلى العمر الزمني، وضرب القيمة الناتجة في ١٠٠، واستخدمت هذه النسبة في المقياس الذي تعديله عام ١٩٠٨.

وقد أدى اختبار بينيه إلى انتشار فكرة قياس الذكاء في العالم وبخاصة في الولايات المتحدة الأمريكية. حيث قام لويس تيرمان Terman وهو أحد علماء النفس

بجامعة ستانفورد عام ١٩١٦ بمراجعة مقياس بينيه وترجمته إلى اللغة الإنجليزية، وأطلق عليه مقياس ستانفورد بينيه للذكاء Stanford - Binet Intelligence Scale .

وقام تيرمان بمراجعة المقياس عام ١٩٣٧ بحيث اشتمل على صيغتين متكافئتين (ل) ، (م) واستخدم مفهوم نسبة الذكاء التى اقترحها شترن، وأعاد مراجعته عام ١٩٦٠، وأصبح صيغة واحدة (ل - م) ضمت أفضل مفردات كل من الصيغتين، وتوسع المقياس بحيث أصبح يقيس مستوى الراشد المتفوق. واستبدل مفهوم نسبة الذكاء (IQ) بمفهوم نسبة الذكاء الانحرافية Deviation IQ (DIQ) التى أشرنا إليها فى الفصل السادس، والتى اقترحها ويكسلر Wechsler عالم النفس الأمريكى فى مقياسه الذى أعده لقياس الذكاء، وهو من الاختبارات الفردية، ويسمى مقياس ويكسلر - بلفيو للذكاء Wechsler - Bellevue Intelligence Scale نسبة إلى مستشفى بلفيو للأمراض النفسية التى كان يعمل بها بمدينة نيويورك.

ونظراً لأن مقياس بينيه يُعد من المقاييس الفردية Individual Test التى تطبق على فرد واحد فى وقت واحد، فقد اتجهت جهود مجموعة من علماء النفس وبخاصة آرثر أوتيس Otis وكان مسئولا عن دور نشر عالمية ، وميلر Miller بجامعة مينيسوتا، وثورنديك Thorndike بجامعة كولومبيا ، وتيرمان Terman، إلى بناء اختبارات جماعية Group Tests .

وقد شجع على ذلك قيام الحرب العالمية الأولى التى تطلبت فرز وتصنيف مئات الآلاف من المجندين فى الجيش فى وقت قصير. لذلك تم إعداد أول اختبار جماعى لقياس الذكاء عام ١٩١٧ ويسمى اختبار الجيش ألفا Army Alpha Test وهو اختبار لفظى. ونظراً لأن بعض المجندين فى ذلك الوقت لم يكن مستواهم فى القراءة جيداً، فقد تم إعداد اختبار جماعى آخر مناسب للأميين ويسمى اختبار الجيش بيتا Army Beta Test وهو اختبار غير لفظى.

وقد اختُبرت أعداد كبيرة بهذين الاختبارين عامى ١٩١٧ ، ١٩١٨ ، وبحلول عام ١٩٢٢ بلغ عدد اختبارات الذكاء الجماعية المنشورة ٤٤ اختباراً.

وعندما قامت الحرب العالمية الثانية انتهز علماء النفس العاملين بالجيش التطور الذى حدث فى بحوث القياس العقلى وأعدّوا اختباراً جماعيًا للذكاء يعرف باسم اختبار التصنيف العام لأفراد القوات المسلحة Army General Classification Test (AGCT) طبق على ملايين الجنود عند التحاقهم بالجيش، وكان يشتمل على أربع صيغ متكافئة يتطلب تطبيق كل منها ساعة واحدة تقريبًا. كما أُعدّت صيغ أطول للاختبار لكى يمكن تقسيم أداء المختبر إلى أربعة أقسام هى : القدرة اللفظية، والفهم المكانى، وإجراء العمليات الحسابية، والاستدلال الحسابى. وقد حل اختبار التأهيل

للقوات المسلحة Army Forces Qualification Test (AFQT) محل هذا الاختبار بعد ذلك.

وقد أسهم في الانتشار السريع لاختبارات الذكاء الجماعية تعميم التعليم وتبنّي الأساليب الكمية في تطوير الإدارة المدرسية في الولايات المتحدة الأمريكية.

فقد أُعدَّت كثير من اختبارات الذكاء الجماعية لجميع الأعمار من الحضانة إلى الجامعة، وأدى ذلك إلى تعميم وتنفيذ برامج اختبارية على نطاق واسع متمثلة في اختبارات القبول بالكليات، واختبارات انتقاء الأفراد وتصنيفهم وتسكينهم، واختبارات الفئات الخاصة وغيرها.

ولعل أهم مصدر من مصادر المعلومات المتعلقة بالاختبارات العقلية هو الكتاب السنوي للقياس العقلي Mental Measurement Yearbook الذي حرَّره أوسكار بيروس Buros عام ١٩٣٨، ولا يزال إصداره مستمرًا حتى وقتنا الحالي. ويقدم هذا الكتاب السنوي قوائم مراجع شاملة تتعلق بالقياس النفسي والتربوي، ومراجعات ناقدة للاختبارات المنشورة والاختبارات الجديدة، والمؤلفات التي تتناول القياس النفسي والتربوي.

فمثلاً الكتاب السنوي السابع الذي صدر عام ١٩٧٢ أشار إلى ١١٥٧ اختبار، ٧٩٨ مراجعات اختبارات بواسطة ٤٣٩ محكم، ١٨١ خلاصة مراجعات، ٩٣ دورية سيكولوجية، ١٢٣٧٢ مصدرًا من مصادر اختبارات سيكولوجية وتربوية.

غير أن هذا الانتشار المتسارع لاختبارات الذكاء لم يواكبه تطورات موازية في أساليب ومنهجيات القياس النفسي بما يسهم في التغلب على العديد من المشكلات وأوجه النقد بل وأحيانًا الهجوم على اختبارات الذكاء، وبخاصة مفهوم نسبة الذكاء كما سيتضح في سياق مناقشتنا لمفهوم الذكاء وكيفية قياسه.

مفهوم الذكاء والمشكلات المتعلقة بتعريفه:

يُعد مفهوم الذكاء من أكثر المفاهيم السيكولوجية التي يدور حولها الحوار والجدل ليس فقط بين علماء النفس والمشتغلين بالقياس وإنما أيضًا بين مستخدمي اختبارات الذكاء والمختبرين بها، بل امتد الجدل خارج النطاق الأكاديمي إلى المسئولين عن اتخاذ القرارات في مختلف المجالات التطبيقية التربوية والنفسية والاجتماعية. ويرجع ذلك إلى ما ترتب على هذا المفهوم وأساليب قياسه من مشكلات متعددة اجتماعية وتربوية، مثل إثارة الضغائن بين أفراد المجتمع الواحد نتيجة تحيز مفردات اختبارات الذكاء ضد مجموعات أقلية تعيش في المجتمع نفسه، والتفرقة

العنصرية بين الأجناس، والتشخيص غير الدقيق للأطفال المتخلفين عقليًا، والانتقاء غير الصائب لأفضل الأفراد للالتحاق بدراسات متقدمة أو القيام بمهام معينة، وغير ذلك من المشكلات. فمفهوم الذكاء يتسم بتعدد تعريفاته وتنوعها نظرًا لعدم وضوح المقصود منه على وجه التحديد مما أدى إلى اختلاف وجهات نظر علماء النفس حوله. وعلى الرغم من أن الجيل الأول ممن اشتغلوا بالقياس انصب اهتمامهم على مشكلة تعريف الذكاء، إلا أنهم لم يتوصلوا إلى تعريف علمى لهذا المفهوم، بل ولم يتفقوا على تعريف يمكن الإفادة منه عمليًا.

غير أن هذا لم يمنع علماء القياس من الاستمرار فى محاولاتهم لقياس الذكاء، وبنائهم اختبارات تتميز درجاتها بالثبات وفاعليتها فى التنبؤ، مما يدل على أن هذه الاختبارات تقيس شيئًا معينًا. وقد أدى هذا الوضع إلى زيادة حدة الجدل حول هذا المفهوم أكثر مما ساهم فى إلقاء الضوء عليه. فمفهوم الذكاء يصعب تعريفه تعريفًا واضحًا ومباشرًا ليس بسبب طبيعة الذكاء، وإنما بسبب طبيعة المفاهيم المجردة بعامة (Neisser 1979) .

وقد أسهم فى هذا الوضع النظر إلى الذكاء على أنه شيء مادى أو كينونة لها وجود فعلى يمكن فحصها، بدلا من النظر إليه على أنه مصطلح يشير إلى سلوك له دلالته فى حياة الإنسان. فالبعض يظن أن المصطلحات أو المفاهيم التى تقترح لوصف الظواهر السلوكية أو غيرها من الظواهر تحمل بين طياتها الشيء الذى يمثلها مما يؤدى إلى الخلط بين مصطلح معين وما يمثله من سلوك أو أفعال معينة، فكلمة «طقس» مثلاً تصف فعل أو سلوك ظاهرة طبيعية معينة، وكلمة «ذكاء» أيضًا تصف ظاهرة سلوكية معينة.

ففى كلتا الحالتين نقيس الفعل أو السلوك ولا نقيس المفهوم ذاته. غير أنه فى حالة «الطقس» يمكن قياس ظواهر متعددة قياسًا مباشرًا، وتشير فى مجملها إلى ما اصطُلح على تسميته «طقس» بينما فى حالة «الذكاء» ربما يكون الأمر أكثر صعوبة. وعادة نصدر فى ضوء هذه القياسات أحكاما تتعلق بالظاهرة أو ما تتضمنها.

فنقول مثلاً أن الطقس جاف أو بارد أو حار، كذلك يمكن القول أن ذكاء الفرد مرتفع أو منخفض أو أن السلوك اللفظى للفرد افضل من سلوكه غير اللفظى، وهكذا. فقياس الشيء أو إمكانية قياسه إذن لا يعنى أن هذا الشيء له وجود مادى.

الذكاء كمفهوم وصفى :

فى ضوء ما تقدم عمل بعض العلماء الذين حاولوا تعريف الذكاء على تجنب التعريفات الصورية، أو الجامدة للذكاء، لذلك لجأوا إلى وصفه أو تحديد خصائصه أو

ما يترتب عليه من سلوك إنساني، وذلك على الرغم من أن بعضاً آخر منهم استخدم مفهوم الذكاء على أنه يعنى الاستعداد الفطرى للفرد مما لا يمكن قياسه قياساً مباشراً. ولعل ما أشار إليه كاتل (Cattell , 1963) حول مفهوم الذكاء المائع Fluid Intelligence الذى اعتبره استعداداً أساسياً للتعلم وحل المشكلات مستقل عن التعليم والخبرة يعبر عن هذا التوجه.

أما الذكاء المتبلور Crystalized Intelligence الذى أشار إليه أيضاً فقد اعتبره ينمو نتيجة تفاعل الذكاء المائع للفرد مع بيئته أو ثقافته، ويشتمل على المعارف والمهارات المتعلمة. ومن أمثلة تعريفات الذكاء في ضوء ما يترتب عليه من سلوك تعريف بينيه Binet وسيمون Simon (١٩٠٥) على أنه تجمع من القوى أو القدرات العقلية، والحكم ، والحس العملي، والمبادأة، والقدرة على التكيف مع الظروف المحيطة. وتعريف تيرمان Terman (١٩٢٠) على أنه القدرة على التفكير المجرد، وتعريف ويكسلر Wechsler (١٩٩٨) على أنه قدرة الفرد الكلية على العمل الهادف، والتفكير المنطقى، والتعامل مع بيئته بفاعلية. غير أن بياجيه Piaget (١٩٨٩) يرى أن الذكاء امتداداً للتكيف البيولوجي ويتضمن عمليات التمثُّل Assimilation وهى العمليات التى تعتمد على مخططات معرفية داخلية تتوافر لدى الفرد، وعمليات المواءمة Accomodation وهى العمليات التى تهدف لتعديل المخططات المعرفية الداخلية لكى تتواءم مع المثيرات البيئية الجديدة التى يواجهها الفرد.

فعمليات التمثل تسهم في الابتعاد عن السلبية فى التعامل مع البيئة، بينما تسهم عمليات المواءمة فى ضبط توجه الذكاء نحو التمثيل المعرفى الذى يتفق مع العالم الواقعى.

وعلى جانب آخر نلاحظ أن Thorndike (١٩٢٨) حاول الابتعاد فى أعماله الرئيسة المتعلقة بقياس الذكاء عن تعريف الذكاء تعريفاً صورياً Formal Definition ، حيث يبدو أنه لم يقترح تعريفاً محدداً فى سياق هذه الأعمال، فقد رأى أن الذكاء يتعلق بمقارنة الفرد بأقرانه من حيث قدرته على إنجاز تسلسل من المهام التى اعتبرها نواتج عقلية.

وفى ضوء هذا الإرباك فى تعريف الذكاء اقترح بيرت Burt (١٩٤٧) أنه من المناسب تذكر التمييز بين التعريف الاسمى Nominal Definition ، والتعريف الواقعى Real Definition.

فالتعريف الاسمى يتعلق بكيفية استخدام المفهوم أو المصطلح، أما التعريف الواقعى فيتعلق بتفسير طبيعة المفهوم المراد تعريفه، ويرى أنه ربما يكون من الضرورى تعريف الذكاء تعريفًا اسميًا قبل التفكير فى كيفية قياسه حيث إن هذا يتطلب مزيدًا من البحث.

ويبدو أن هذا الاقتراح كان له صدى واسعًا بعد ذلك فى الأوساط السيكولوجية، والدليل على ذلك تعدد المؤتمرات المتخصصة التى انعقدت فى الولايات المتحدة الأمريكية وكندا لمناقشة مجال الذكاء الإنسانى. ومثال ذلك مؤتمر تورنتو بكندا «حول الذكاء On Intelligeence » عام ١٩٦٩، وسلسلة المؤتمرات التى عقدها مركز بحوث وتطوير التعلم بجامعة بيتسبرج بالولايات المتحدة الأمريكية المتعلقة بمجالات مهمة فى سيكلوجية التعليم، وكان أحد هذه المجالات « مجال الذكاء وطبيعته The Nature of Intelligence» وذلك عام ١٩٧٤ وأسهم فيه نخبة من علماء القياس واختصاصيى علم النفس المعرفى، والندوات المتعددة التى نظمتها الجمعية الأمريكية للبحث التربوى من خلال اجتماعها السنوى، مثل الندوة التى عقدت فى تورنتو بكندا عام ١٩٧٨، وندوة سان فرانسيسكو عام ١٩٧٩ حول الذكاء الإنسانى Human Intelligence، وكذلك إصدار دورية الذكاء Intelligence وتُعد من الدوريات السيكولوجية المتخصصة التى تهتم بنشر البحوث والدراسات المبتكرة التى تسهم فى إثراء الفكر فى مجال الذكاء الإنسانى من حيث طبيعته ووظائفه وقياسه.

وقد اهتمت هذه المؤتمرات اهتمامًا خاصًا باقتراح التوجهات المستقبلية لمفهوم الذكاء الإنسانى ، وأساليب قياسه من منظور علم النفس المعرفى Cognitive Psychology، وقد تبنت بعض الدراسات التى قُدمت فى هذه المؤتمرات مفهوم «الشخص الذكى Intelligent Person» بدلاً من مفهوم «الذكاء Intelligence» حيث اعتبرت أن هذا الشخص يعد مثالاً نموذجيًا تصوريًا Prototype يُقارن به الأشخاص الآخرين لمعرفة مدى مماثلتهم له، وتعتمد ثقتنا فى الحكم بأن أحد الأشخاص الآخرين «ذكى» على مماثلته بوجه عام لهذا المثال النموذجى، كما هو الحال عندما نطلق على أحد الأشياء اسمًا معينًا مثل «مقعد»، فثقتنا فى هذه التسمية تعتمد على مماثلة هذا الشيء للمثال النموذجى للمقعد. وتفضل روش Rosch وزملاؤها استخدام لفظ «قسم Category» فى تحليل بنية المفاهيم Concepts

(Rosch ,1978 , Rosch , Mervis , Gray , Johnson & Boyes , 1976).

وينطبق هذا أيضًا على مفهوم «الشخص الذكى» . وعلى الرغم من أن هؤلاء الباحثين يرون أنه لا توجد مجموعة معينة من الخصائص تُعرِّف هذا «القسم» تعريفًا دقيقًا، إلا أنه يجب أن تكون هناك خصائص يشترك فيها أعضاء هذا «القسم» لكى يُطلق عليه «قسمًا Category» .

ولعل «المثال النموذجي Prototype» للشخص الذكي يُعد أحد العناصر التي تنتمي إلى هذا القسم إذا توافرت فيه جميع الخصائص المشتركة، غير أنه لا توجد دراسات علمية منظمة لهذه الخصائص، ولكن يمكن تتبع بعض الدراسات الاستطلاعية في هذا الشأن مثل دراسة ديوب (Dube, 1977) وغيرها ، حيث بينت عدم الاتفاق على هذه الخصائص. ولعل هذا يرجع إلى عدم وجود محك محدد للذكاء نظرًا لتعدد مظاهره، إذ يمكن أن يتميز شخصان بالذكاء على الرغم من اختلافهما في كثير من السمات، أي أنهما يمثلان «المثال النموذجي التصوري للشخص الذكي» في أبعاد مختلفة.

لذلك يرى نايسر (Neisser, 1979) أن ما نُطلق عليه «مقعد Chair» لمماثلته لنموذج مقعد مثالي تصوري Prototype لا يعني أن هناك خاصة تسمى «مقعدية Chairness» ، كذلك ما نطلق عليه «شخص ذكي» لمماثلته «النموذج شخص ذكي مثالي تصوري» لا يعني أن هناك شيئًا ما يسمى «ذكاء Intelligence» .

فالمماثلة تعد حقيقة خارجية وليست مضمونًا أو جوهرًا داخليًا، وبذلك لا نستطيع تعريف الذكاء تعريفًا عملياتيا Process - Based نظرًا لأنه متعدد الخصائص، ولكنه يعد مماثلة بين شخصين أحدهما شخص فعلي والآخر نموذج مثالي تصوري، وهذه المماثلة هي التي نعمل على قياسها في أبعاد متعددة.

ومن الجدير بالذكر أن ثورنديك Thorndike (١٩٢٤) باعتباره أحد الرواد الأوائل المرموقين في القياس النفسي كان قد اقترح منظورًا مماثلاً لهذا المنظور الجديد لمفهوم الذكاء.

ولعل هذا المنظور يلفت النظر إلى أننا لا نستطيع تعريف الذكاء تعريفًا لفظيًا، ومع هذا يمكن قياسه وذلك بضم قياسات جميع المكونات المناسبة بأوزان اختيارية لكل مكونة منها للحصول على مؤشر عام.

غير أن كثيرًا من هذه المكونات لا نستطيع قياسها بأي أسلوب مقنن، فخصائص المثال النموذجي للشخص الذكي لا تقتصر على الطلاقة اللفظية، والقدرة المنطقية، والمعلومات العامة، وإنما تتضمن أيضًا الفطنة، والذوق العام، والتفكير الابتكاري، وعدم التحيز، والاستقلال الفكري، والتفتح العقلي، وغير ذلك.

وبعض هذه الخصائص تتمثل في مواقف حياتية مميزة، والبعض الآخر لا نستطيع تقويمه إلا في ضوء حياة الفرد ككل، ومع هذا فإن علماء القياس يسعون إلى قياس الخصائص الأكثر تعلقًا بالمثال النموذجي للشخص الذكي(Neisser , 1979) .

الذكاء كمجموعة من القدرات المترابطة :

لم يغفل تاريخ القياس النفسى والتربوى البحث عن مؤشرات كمية مقبولة للذكاء الإنسانى، فمعظم علماء النفس الرواد فى مجال الذكاء وقياسه اعتبروا الذكاء مكون من قدرة عامة وقدرات خاصة متعددة.

فنسبة الذكاء IQ كانت وسيلتهم فى تقدير كمّ هذه القدرات وتلخيصه بقيمة عددية واحدة للفرد تمثل ذكاءه، كما هو الحال عند تمثيل توزيع مجموعة من الدرجات بقيمة عددية واحدة تعبّر عن متوسطها الحسابى. غير أن هؤلاء العلماء لم يتفقوا على القدرات الأساسية التى تمثل الخصائص التكوينية للذكاء. ولتوضيح ذلك نعرض فيما يلى بإيجاز بعض التوجهات التى كانت لها أهمية تاريخية ويتناولها الباحثون فى وقتنا الحاضر بالدراسة.

الذكاء كعامل عام :

إن أول محاولة علمية لوصف العناصر التى تشترك فيها جميع اختبارات الذكاء استناداً إلى الأساليب الإحصائية لدرجات هذه الاختبارات قام بها شارلز سبيرمان Spearman عام ١٩٠٤، وأعاد المحاولة عام ١٩٢٧.

ولعله كان من أوائل الذين استخدموا أسلوب التحليل العاملى Factor Analysis الذى أشرنا إليه فى الفصل الخامس فى تحليل البيانات التى استمدها من تطبيق اختبارات الذكاء على تلاميذ المدارس الإنجليزية.

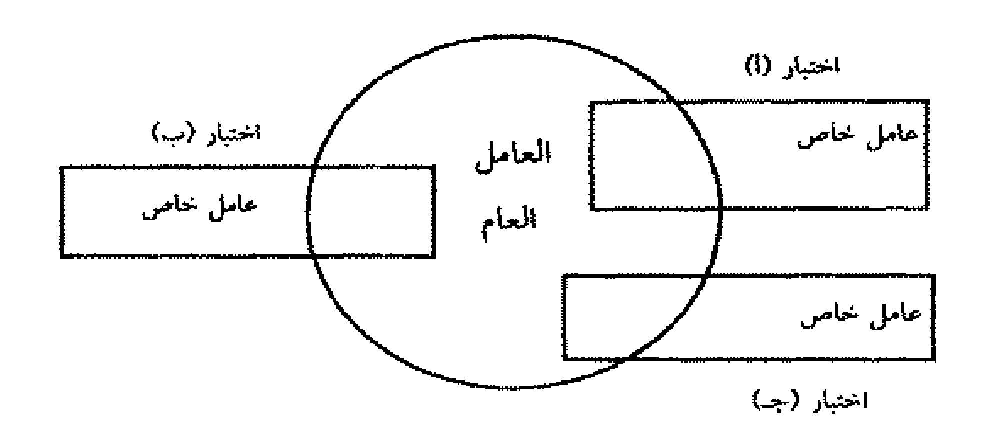
وقد توصل إلى أن جميع مظاهر النشاط العقلى كما تُقاس باختبارات الذكاء تشترك فى عامل عام General Factor (g) عرّفه بأنه القدرة العامة على استنباط العلاقات المجردة. وعلى الرغم من أن بعض المهام المعرفية تتطلب هذه القدرة العامة أكثر من غيرها من المهام، إلا أن جميع هذه المهام تتضمن هذا العامل العام إلى حد ما.

غير أن سبيرمان لاحظ أنه إذا كان العامل العام(g) بمفرده يفسر الذكاء، فإنه من المتوقع أن يكون معامل الارتباط بين اختبارات الذكاء المختلفة تامًا، أى أن جميع قيمه تساوى الواحد الصحيح. غير أنه وجد أن بعض هذه القيم أقل من ذلك بكثير، مما أدى به إلى افتراض وجود عوامل أخرى تسهم فى هذا التفسير أطلق عليها العوامل النوعية Specific Factors (S).

فكل اختبار للذكاء يقيس نشاطًا عقليًا يمكن تحليله إلى عاملين أحدهما عام (g) مشترك بين هذا النشاط أو محتواه، وهو ما أسماه الذكاء العام General Intelligence وعوامل خاصة (S) تتعلق بهذا النشاط دون غيره من الأنشطة.

فالعامل العام (g) إذن يسهم في الارتباط بين اختبارات الذكاء، بينما تعمل العوامل الخاصة على تحجيم الارتباط حتى لا تصل قيمته إلى الواحد الصحيح.

لذلك فإن اختبارات الذكاء التي تُبنى في ضوء هذا التوجُّه ينتج عنها درجة كلية واحدة للذكاء بحيث تعبِّر بقدر الإمكان عن العامل العام دون الاهتمام بالعوامل الخاصة نظرا لأنها غير مشتركة بين الاختبارات المختلفة. ويوضح الشكل التخطيطي (٩ ـ ١) نظرية العامل العام لسبيرمان:



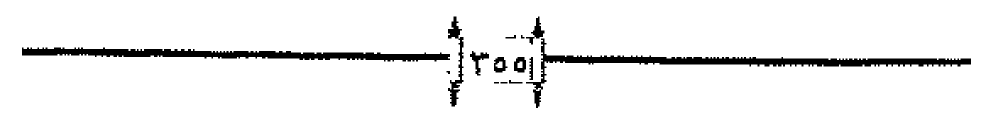
شكل (٩ ـ ١) يوضح نظرية العامل العام لسبيرمان

ويتضح من شكل (٩ ـ ١) أن كل مستطيل يُمثل اختباراً يقيس العامل العام (g) بقدر الإمكان. والمساحة الناتجة من تقاطع كل مستطيل مع الدائرة التي تُمثل العامل العام تدل على مقدار ما يقيسه الاختبار من العامل العام. أما أجزاء المستطيلات التي تقع خارج الدائرة فإنها تُمثل العامل الخاص بكل اختبار.

ومن الجدير بالذكر أن سبيرمان أجرى تعديلاً على نظرية العامل العام بحيث تأخذ بعين الاعتبار العوامل التي لم تفسَّر بواسطة العامل العام أو العوامل الخاصة، فقد وجد أن هناك بعض المهام العقلية الخاصة ولكنها تكون مرتبطة فيما بينها. لذلك أطلق عليها «العوامل الطائفية Group Factors »، وهذه العوامل أقل اتساعًا من العامل العام وأقل تخصصًا من العوامل الخاصة، أي أنها تتوسط بين كل من نوعي العوامل.

ولعل هذه النظرية قد أسهمت في الفكر التربوي في ذلك الوقت عندما كان يُفترض أن التلميذ الذكي أو الموهوب يكون كذلك في جميع المجالات الدراسية دون تخصيص.

وقد قوبلت هذه النظرية بعد ذلك بهجوم شديد من جانب علماء النفس في الولايات المتحدة الأمريكية بحجة أن هذه النظرية قد اعتمدت في تحليلاتها الإحصائية



على عدد ضئيل من التلاميذ الذين طُبِّقت عليهم الاختبارات، وهذه العينة لا تصلح أساسًا للتوصل إلى ما أطلق عليه سبيرمان «العامل العام» . كما أن الأسلوب الإحصائى الذى اتبعه سبيرمان لا يوضح الأوزان النسبية للعوامل التى تتدخل فى كل اختبار.

وقد اعترض جيلفورد (Guilford, 1976) على فكرة أن جميع اختبارات الذكاء ترتبط درجاتها ارتباطًا موجبًا، حيث أكد أن البيانات الإمبيريقية التى جمعها لم تؤيد ذلك. وأشار إلى مئات من الارتباطات الصفرية بين درجات هذه الاختبارات المستمدة من عينات بلغ عدد أفراد كل منها ما يقرب من ٢٢٥ فردًا. كما أشار إلى أن قيم الارتباطات الموجبة التى حصل عليها سبيرمان ربما ترجع إلى استخدامه عددًا قليلاً من الاختبارات المحدودة فى محتواها العاملى. وأهم من ذلك أن هذه الاختبارات يبدو أن تكوينها العاملى مركب مما أدى إلى التداخل فيما بينها.

ويرى جيلفورد أن التوصل إلى ما يسمى «العامل العام» (g) يتطلب عينات غير متجانسة من الأفراد ، نظرًا لأن هذا العامل ربما يمثل التباين فى العمر الزمنى أو النوع أو التعليم وما إلى ذلك. فإذا ارتبطت درجات اختبارين بالعمر الزمنى فإن الاختبارين سوف يرتبطان ببعضهما البعض نتيجة لذلك.

الذكاء كعوامل متعددة :

ينظر علماء النفس الأمريكيون إلى الذكاء نظرة مختلفة. فمثلاً ربما يكون فرد لديه قدرة لفظية مرتفعة، وقدرة مكانية وأخرى عددية منخفضة. ويعتمد عدد القدرات الأساسية التى تُفترض فى هذه الحالة على الأساليب الإحصائية المستخدمة لاستخلاص هذه القدرات التى تسمى «عوامل Factors» ، كما يعتمد عددها على الإطار النظرى الذى يتبناه الباحث فى القدرات العقلية.

ولعل لويس ثيرستون Thurstone أحد علماء النفس البارزين بجامعة شيكاغو الأمريكية (١٩٣٨) هو الذى اهتم اهتمامًا ملحوظًا بمدخل العوامل المتعددة فى الذكاء. حيث تركزت كثير من دراساته وبحوثه فى هذا المجال. فقد امتد نشاطه إلى تطوير أساليب ومناهج البحث فى مشكلة التكوين العقلى وبخاصة منهج التحليل العاملى.

فقد سبق أن ذكرنا أن التحليل العاملى يُعد طريقة منظمة لفحص مكونات الاختبارات وارتباطها ببعضها البعض. فبعد تطبيق مجموعة من الاختبارات على عينة من الأفراد نحاول باستخدام منهج التحليل العاملى تحديد أقل عدد من العوامل أو القدرات المميزة التى يمكن قياسها بدرجة متسقة والتى تكمن وراء أداء أفراد العينة فى مجموعة الاختبارات.



وقد توصل ثيرستون إلى مجموعة من القدرات العقلية الأولية Primary Mental Abilities نتيجة تطبيق منهج التحليل العاملى على ٦٥ اختباراً طبقها على مجموعة من الطلاب فى مرحلة التعليم الثانوى، واستطاع تحديد مجموعة من العوامل المستقلة نسبيًا عن بعضها البعض الآخر ، وإعداد اختبار لقياس هذه العوامل يسمى اختبار القدرات العقلية الأولية.

وهذه العوامل هى :

(١) القدرة على الطلاقة اللفظية : القدرة على استخدام الكلمات فى مواقف مختلفة.

(٢) القدرة على الفهم اللغوى : فهم معانى الكلمات والعلاقات القائمة بينها فى الجمل.

(٣) القدرة العددية: القدرة على إجراء العمليات الحسابية البسيطة بسرعة ودقة.

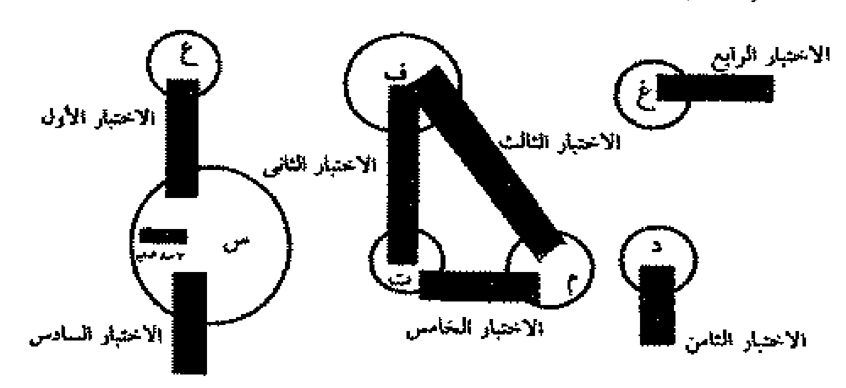
(٤) القدرة المكانية : القدرة على إدراك العلاقات الهندسية والمكانية الثابتة والتى يجرى عليها تحويلات معينة.

(٥) السرعة الإدراكية : الملاحظة السريعة والدقيقة لتفاصيل الأشياء المنظورة.

(٦) القدرة على التذكر : القدرة على الاسترجاع الفورى لمثيرات محددة.

(٧) القدرة على الاستدلال الاستقرائى : ويقصد به الاستنباط العام، أى القدرة على استخلاص القواعد أو المبادئ.

ويمكن تمثيل نظرية ثيرستون أو نموذجه متعدد العوامل بالشكل التخطيطى (٩_٢) التالى (Sax, 1969) :



شكل (٩ ـ ٢) يوضح نظرية العوامل المتعددة لثيرستون

ويتضح من شكل (٩ ـ ٢) أن كل دائرة تُمثل عاملا مستقلا استقلالاً تامًا عن غيره من العوامل. فالاختبار الأول يقيس بدرجة ضئيلة العامل العددى (ع)، وعامل

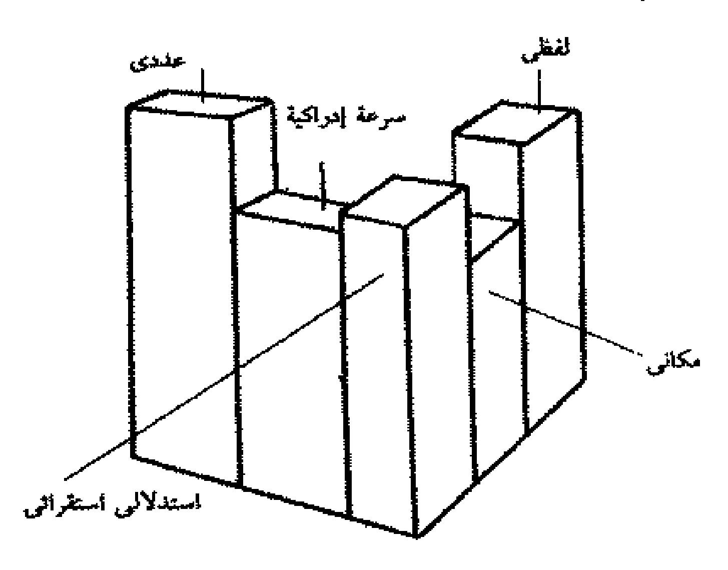
الاستدلال الاستقرائى (س)، والاختبار الثانى يقيس بدرجة كبيرة كل من العامل اللفظى (ف)، وعامل التـذكر (ت) ، ولكنه لا يصلح لقيـاس العامل المكانى (م) ، والاختـبار الثالث يقيس بدرجة أكبر العامل اللفظى وبدرجة أقل العامل المكانى، أما الاختبار الرابع فلا يصلح إلا لقياس العامل اللغوى (غ)، والاختبار الخامس يقيس كل من عامل التذكر (ت) والعامل المكانى (م)، أما الاختباران السادس والسابع يُعدان مقياسين جيدين لعامل الاستدلال الاستقرائى (س)، بينما يُعد الاختبار الثامن مـقياسًا نقيًا نسبيًا لعامل السرعة الإدراكية (د).

وقد أعـيد تقنين وتحليل القـدرات العقليـة الأولية عام ١٩٦٢ وتم الـتوصل إلى خمسة عوامل هى:

العامل اللفظى، والعامل العددى، وعامل الاستدلال الاستقرائى، وعامل السرعة الإدراكية، والعامل المكانى.

غيـر أن العوامل الثـلاثة الأولى استُـخلصت من نتائج تطبـيق الاختـبارات على تلاميذ الصـفوف من الرابع الابتدائى إلى الثانوية العامـة، أما العامل الرابع فقد اعـتمد على نتائج تطبيق الاختبارات على الصفوف من الحضانة إلى الصف السادس.

ويوضح الشكل التخطيطى (٩ ـ ٣) التالى هذه العوامل:



شكل (٩ ـ ٣) القدرات العقلية الأولية بعد إعادة تقنين وتحليل اختبار ثيرستون

وقد حاول ثيرستون فيما بعد تحليل العلاقات بين العوامل الأولية التى استخلصها لكى يتوصل إلى ما أسماه العوامل العامة من الدرجة الثانية - Second Order General Factors بغرض تفسير الارتباطات بين القدرات العقلية الأولية التى كانت قيمها متوسطة حيث وجد أن وسيط هذه القيم ٠,٣٥ . وعلى الرغم من أن نظرية ثيرستون استبعدت فى البداية العامل العام (g) الذى اقترحه سبيرمان كمكونة لها دلالتها فى الوظائف العقلية، إلا أن الارتباطات بين القدرات العقلية الأولية التى وجدها ثيرستون فى تحليله الأخير جعله يفترض وجود عامل من الدرجة الثانية ربما يكون له علاقة بالعامل العام (g) .

وبالطبع يمكن استخلاص «عوامل Factors» أخرى إذا أضفنا مهامًا عقلية مختلفة إلى ما تتضمنه الاختبارات التى استخدمها ثيرستون وتطبيقها على عينات أخرى من الأفراد من ذوى خلفيات ثقافية متباينة.

وعمومًا فإن استخدام نظرية العوامل المتعددة لثيرستون لتقييم الذكاء يتطلب تطبيق بطارية من الاختبارات البسيطة التى يصمم كل منها لتقييم عامل خاص أو أولى من العوامل السابقة، مثل اختبار القدرات العقلية الأولية لثيرستون ، وبطارية الاستعدادات الفارقة التى سوف نوضحها فى الفصل العاشر.

الذكاء كبنية ثلاثية الأبعاد :

توجد محاولات أخرى من جانب علماء القياس النفسى لتصنيف العوامل التى استخلصتها البحوث المختلفة نتيجة استخدام منهج التحليل العاملى، ومن أهم الجهود فى هذا المجال بحوث جليفورد Guilford وزملاؤه بجامعة جنوب كاليفورنيا التى تم توثيقها فى سلسلة من التقارير بعنوان عام «تقارير من معمل علم النفس».

وتتضمن هذه السلسلة أعمال جليفورد وزملاؤه، والأساليب والمنهجيات والبحوث التى أنجزها الفريق البحثى على مدى عقدين من الزمان. وفى كتابه الذى نشر عام ١٩٦٧ بعنوان «طبيعة الذكاء الإنسانى» اقترح جليفورد نموذجه المعروف «بنية العقل (Structure of Intellect (SOI» الذى يسهم فى الوظائف الكشفية التنقيبية Hueristic Model لتكوين الفروض المتعلقة بعوامل جديدة للذكاء. لذلك افترض لهذه البنية نظامًا ثلاثى الأبعاد أو الأوجه Facets يتضمن العديد من القدرات العقلية التى صُنفت استنادًا إلى هذا النظام كالتالى:

(١) المحتوى Content: أى المادة التى يجرى تناولها.

(٢) العمليات Processes أو الإجراءات Procedures التى تُؤدَّى على المحتوى.

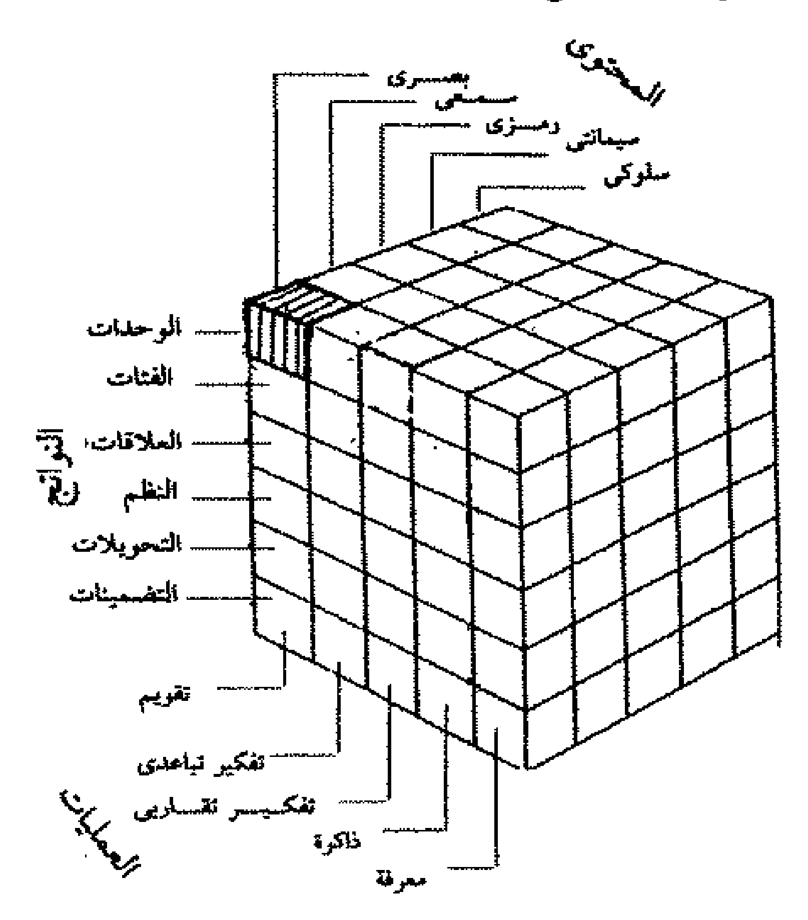


(٣) النواتج Products : أى ما ينتج عن أداء العمليات على المحتوى، أو صيغة الفكر الناتج.

وقد قسَّم كل من هذه الأبعاد أو الأوجه بعد تحليلها إلى أقسام فرعية، أربعة للمحتوى، وخمسة للعمليات، وستة للنواتج، وبذلك يصبح العدد الكلى للأقسام :

٤ × ٥ × ٦ = ١٢٠ قسمًا . غير أن جيلفورد عام ١٩٧٠ أعاد النظر فى البُعد المتعلق بالمحتوى حيث قسَّم أحد مكوناته وهو المحتوى الشكلى إلى محتوى بصرى، ومحتوى سمعى ، وبذلك أصبح عدد أقسام المحتوى خمسة ، والعدد الكلى للأقسام :

٥ × ٥ × ٦ = ١٥٠ قسمًا موضحة بالشكل التخطيطى (٩ ـ ٤) التالى:



شكل (٩ ـ ٤) يوضح نموذج جيلفورد ثلاثى الأبعاد أو الأوجه

ويمكن توضيح مكونات الأبعاد الثلاثة التى يمثلها شكل (٩ ـ ٤) كالتالى:

(١) **المحتوى** Content : ويعتمد أساسًا على التمييز بين الأنواع المختلفة من

المعلومات التى يمكن أن تتضمنها مشكلة أو مهمة معينة يجابهها الفرد ويعمل عليها العقل.

وينقسم المحتوى فى ضوء ذلك إلى خمسة أقسام هى :

المحتوى السلوكي Behavioral :ويتضمن معلومات تتعلق بسلوك الأفراد والتفاعل فيما بينهم، ويكون السلوك أساسًا غير لفظي حيث يتركز الاهتمام على الوعى أو الانتباه، والإدراكات ، والرغبات، والمشاعر، والأمزجة، والانفعالات، والتصرفات. وقد أضيف هذا القسم على أساس نظرى بحت ليمثل مجالا عامًا يُطلق عليه أحيانًا «الذكاء الاجتماعي».

المحتوى السيمانتى Semantic : يتضمن معانى لفظية وأفكار.

المحتوى الرمزى Symbolic : يتضمن معلومات رمزية مثل الحروف والأرقام والإشارات المتعارف عليها.

المحتوى الشكلى Figural: يتضمن مواد عيانية يمكن إدراكها من خلال الحواس، وينقسم إلى محتوى بصرى ومحتوى سمعى.

٢) العمليات processes : هى أساليب النشاط العقلى التى يفترض إجراؤها على محتوى أو معلومات معينة. وتنقسم العمليات إلى خمسة أقسام هى:

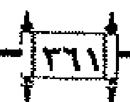
الإدراك المعرفى Cognition : يتعلق بالأنشطة العقلية المعرفية، مثل الفهم والاستيعاب، واكتشاف المعلومات بمختلف أنواعها، وإعادة اكتشافها أو التعرف عليها.

الذاكرة Memory : تتعلق بالاحتفاظ بالمعلومات المتعلمة بحيث يمكن استدعائها بشكلها الأصلى عند الحاجة إليها.

التفكير التقاربي Convergent Thinking : يتعلق بالبحث عن حل صحيح لمشكلة معينة، أو التوصل إلى معلومات محددة من معلومات أخرى معطاه، ويكون التأكيد هنا على المعلومات المتعارف عليها أو التى تحوز قبولاً، وربما تحدد المعلومات المعطاة الاستجابة تحديدًا كاملاً.

التفكير التباعدى Divergent Thinking: يتعلق بالبحث عن حلول ابتكارية متعددة تتميز بعنصر الجدة على الرغم من وحدة مصدر المعلومات.

التقويم Evaluation يتعلق بإصدار أحكام قيمية على المعرفة والفكر من حيث مدى صلاحية المعلومات وملاءمتها وكفايتها والحاجة إليها وما إلى ذلك.



(٣) النواتج Products : هى نتائج أداء العمليات على المحتوى، أى شكل الفكر الناتج. وتنقسم هذه النواتج إلى ستة أقسام هى :

الوحدات Units : تتعلق بإنتاج كلمة أو تعريف أو معلومات بسيطة متمايزة.

الفئات Classes : تتعلق بإنتاج وحدات تشترك فى بعض الخصائص مثل المفاهيم، وملاحظة أوجه التشابه بين الوحدات.

العلاقات Relations : تتعلق بإنتاج أشكال مختلفة من العلاقات مثل أوجه الاتفاق أو الاختلاف أو التشابه.

النظم Systems : وتتعلق بإنتاج مجموعة من التصنيفات متعددة الأشكال أو المحتوى بحيث تتميز بالاتساق الداخلى، مثل نظام بناء الجمل أو نظام العمليات الرياضية.

التحويلات Transformations : تتعلق بتغيير المعانى والتنظيمات والترتيبات وتحويل المعلومات من صيغة إلى أخرى.

التضمينات Implications تتعلق بإنتاج معلومات تتخطى حدود البيانات المعطاة مثل التوقع أو التنبؤ المستقبلى.

ويتضح من شكل (٩ ـ ٤) أن إسهامات جيلفورد الرئيسة تتعلق بتطوير نظام تصنيفى Taxonomy لعوامل الذكاء الإنسانى فى ثلاثة أبعاد أو أوجه لمكعب يشتمل على ١٥٠ خلية تمثل كل منها عاملاً مركبًا تركيبيًا فريدًا من محتوى، وعملية، وناتج. فمثلاً الخلية الأولى المظللة تمثل عامل معرفة وحدات الأشكال، وهكذا فى بقية الخلايا، والوجه العلوى للمكعب يمثل طبيعة المحتوى أو المادة التى تتضمنها مفردات الاختبار.

فالمفردات التى محتواها شكلى تتناول مثلاً تشابه الأشكال، والعلاقات المكانية، والتصور البصرى المكانى، والغلق الإدراكى. والمفردات التى محتواها رمزى تتناول الكلمات المختلطة وغير المختلطة ، وتقيس القدرة على إنتاج أكبر من عدد من الكلمات التى تبدأ بحرف معين، كما تتناول الأعداد والعمليات الحسابية. والمفردات التى محتواها سيمانتى تتناول المعانى اللفظية مثل المتشابهات اللفظية. والمفردات التى محتواها سلوكى تتناول تعبيرات الوجه، والعلاقات الاجتماعية، والتفاعل الاجتماعى، وتضمينات تغيير صورة وجه شخص بصورة أخرى مثلاً، وما يترتب على ذلك من تغيير المعانى نتيجة تغيير الشخصيات والمشاعر والمقاصد والأفكار، وكذلك التنبؤ من

موقف تمثيلي معين بموقف آخر، استناداً إلى مشاعر ومقاصد شخصيات الموقف التمثيلي، وهكذا.

وقد أعد جيلفورد وزملاؤه اختبارات تقيس كثيراً من هذه العوامل التي عددها ١٥٠ عاملاً. فحتى عام ١٩٧١ كان عدد العوامل التي أمكنهم قياسها ٩٨ عاملا (Guilford & Hoepfner , 1961).

وعلى الرغم من أن كثيراً من علماء النفس أشاروا إلى وجود ارتباطات مرتفعة بين الخلايا أي بين العوامل المختلفة، إلا أن النموذج قد استثار الفكر السيكولوجي والتربوي، والدليل على ذلك كثرة البحوث التي أجريت حول كل من النموذج وسلسلة الاختبارات التي صممت لقياس كثير من العوامل التي يتضمنها.

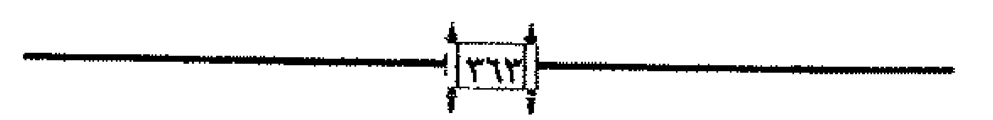
وقد أكد النموذج فكرة الإنسان المتفاعل ، كما أكد مفهوم «المعلومات Information» التي يعمل الإنسان على التمييز بينها في ضوء أنواع مختلفة من النواتج، مثل الوحدات والعلاقات والنظم. فالأقسام الخمسة للمحتوى عندما يتم ضمها في جميع الطرق الممكنة لأقسام النواتج الستة تؤدي إلى ٣٠ قسمًا فرعيًا من المعلومات.

فهذا التصنيف المنظم يمكن النظر إليه على أنه نظرية سيكولوجية للمعرفة تتضمن ٣٠ قسمًا فريدًا تُعرِّف الأنواع الرئيسة للمعلومات. وهذا التصنيف للمعلومات لا يفيد فقط في إثراء النظرية السيكولوجية وإنما أيضًا له تطبيقاته في مجال التربية وتقنيات الحاسوب.

إذ يمكن اعتبار الإنسان كالحاسوب في عمليات تجهيز ومعالجة المعلومات Information Processing ، وهذا التجهيز يستند إلى الأنواع الخمسة من العمليات التي أشار إليها النموذج. غير أن فعل التجهيز Processing Action يختلف باختلاف نوع المعلومات التي يتم تجهيزها أو تناولها .

وهذا التفاعل بين نوع العملية ونوع المعلومات يُعد مسئولا عن التمايز بين القدرات التي تمثلها خلايا المكعب الذي يمثل النموذج (Guilford, 1967). والتشابه بين الإنسان والحاسوب، أو إمكانية استخدام نموذج بنية العقل (SOI) في تقنيات الحاسوب ليست أفكاراً خيالية، فقد اقترح كريستنسن (Christensen, 1963) الاسترشاد بأقسام هذا النموذج في توضيح وظائف الحاسوب، وأشار إلى التوازي بين بعض العوامل العقلية وبعض برامج الحاسوب.

أما في المجال التربوي، فقد أشارت كثير من الدراسات إلي تطبيقات متعددة لهذا النموذج، حيث استخدمت ميكر (Meeker, 1964) النموذج في تقويم كل من

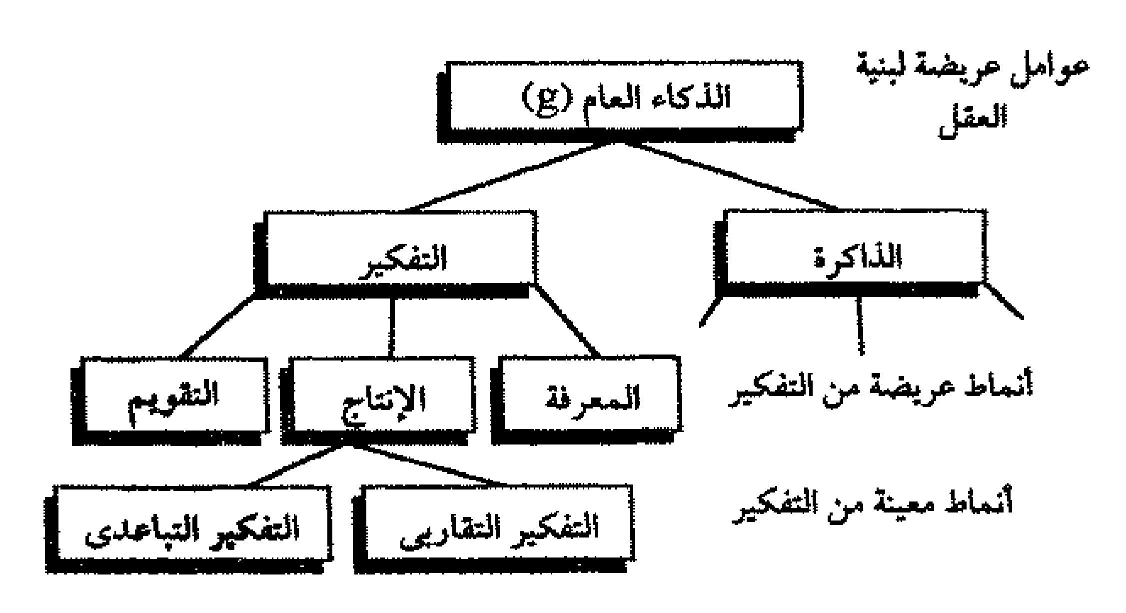


المنهج وقدرات الأطفال، وبينت أنه عند الحديث عن الذكاء يجب على المربين تحديد مجال معين. فعندما يكون الطفل متوسط القدرة ويُخفق في التعلم فإن نسبة ذكائه (IQ) لا تفسر هذا الإخفاق ولا تبرره. وترى أننا نحتاج إلى اختبارات تستند إلى أسس نظرية تزود المعلم بمعلومات تشخيصية تساعده في اتخاذ قرارات تعليمية تتعلق بطفل معين.

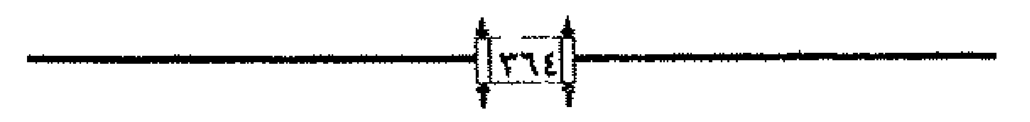
لذلك فإن نموذج جيلفورد يُعد أحد طرق تنظيم القدرات في المناهج المدرسية بدلا من تركيز هذه المناهج على بعض القدرات وبخاصة قدرات الذاكرة على حساب قدرات التفكير وبخاصة التفكير التباعدى والتقويم، وكذلك تركيزه على المحتوى السيمانتى والمحتوى الرمزى، على حساب المحتوى الشكلى، والمحتوى السلوكى، وتركيزه على نواتج الوحدات، والفئات، على حساب نواتج العلاقات، والنظم، والتحويلات، والتضمينات.

وعلى الرغم من أن أبعاد أو أوجه النموذج الثلاثة لم يتأكد انتظامها فى نسق هرمى Hierarchical إلا أن جيلفورد قارن نموذجه بتصنيف بلوم Bloom's Taxonomy ولاحظ أوجه شبه كثيرة بينهما مما جعله يقترح استخدام نموذجه فى تصنيف الأهداف التربوية، وذلك لأنه أكثر دقة ويسمح بتحديد هذه الأهداف تحديدًا إجرائيًا.

ويوضح الشكل التخطيطى (٩ ـ ٥) التالى اقتراح جيلفورد فيما يتعلق بالنسق الهرمى لنموذجه (Guilford, 1959) :



شكل (٩ ـ ٥) نموذج هرمى لعوامل بنية العقل لجيلفورد

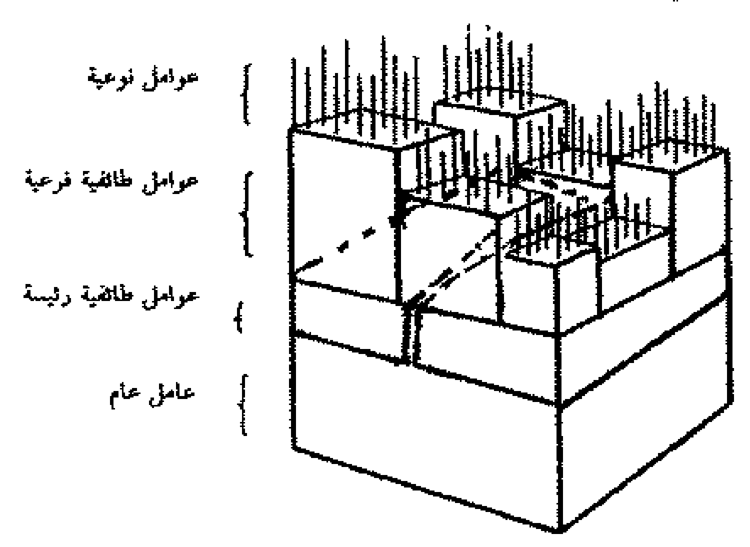


الذكاء كنظام هرمى من العوامل :

اتضح لنا أن نموذج بنية العقل لجيلفورد قدَّم أساسًا نظريًا لطبيعة العقل وأدى إلى تصنيف عدد كبير من العوامل المنفصلة، ولكنه لم يبين العلاقات القائمة بين هذه العوامل مما يتطلب التوصل إلى نموذج رياضى يُستند إليه فى كشف هذه العلاقات أو توضيحها.

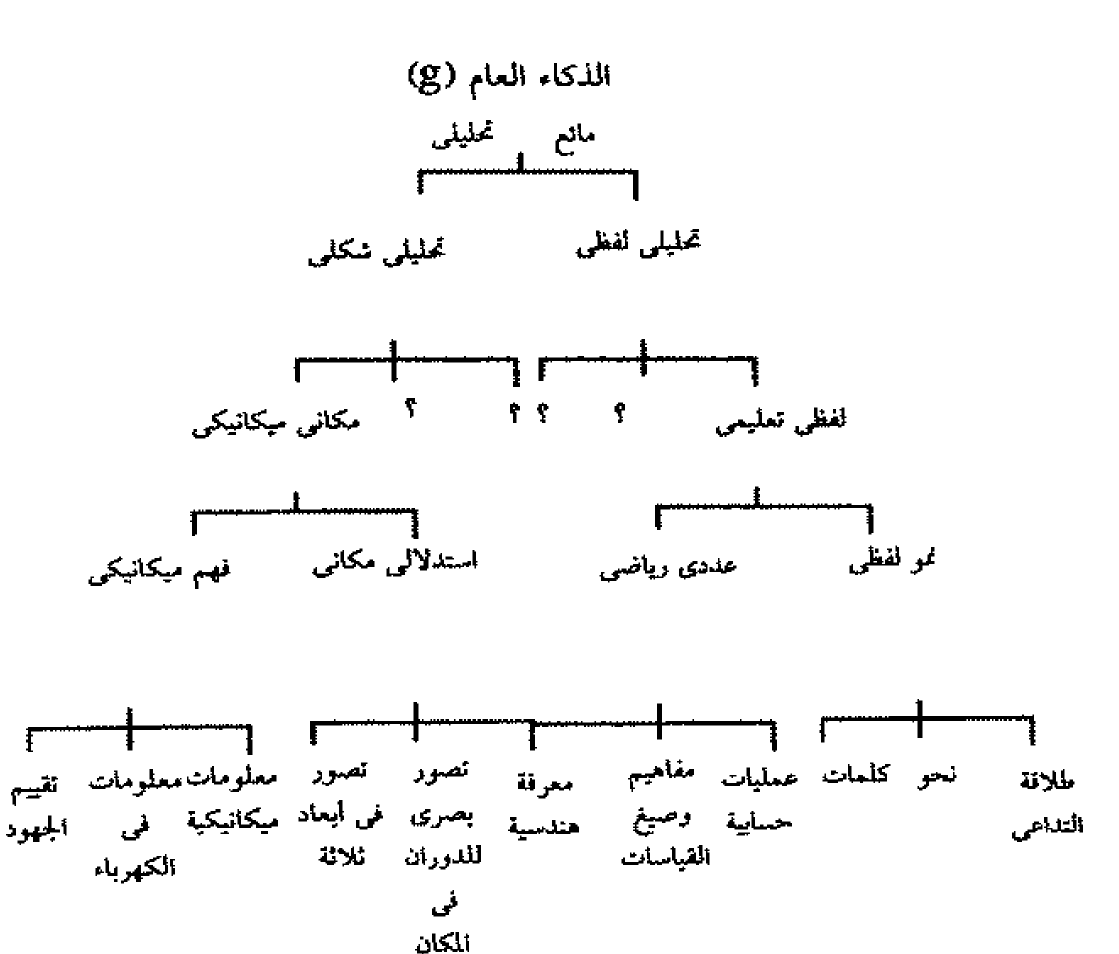
لذلك اقترح كثير من الباحثين استخدام النماذج الهرمية Hierarchical Models التى تمثل مستويات من العوامل متدرجة فى عموميتها. ولعل جيلفورد كان أحد هؤلاء الباحثين حيث اقترح النموذج الهرمى لعوامل بنية العقل الذى أشرنا إليه فيما سبق، والموضح بالشكل التخطيطى (٩ ـ ٥).

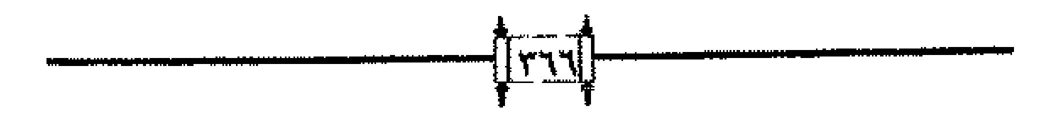
وقد بيَّن فيرنون (Vernon, 1950) عالم النفس الإنجليزى أنه يمكن وصف القدرات العقلية وتصنيفها بوضوح أكثر وذلك بتنظيمها تنظيمًا هرميًا فى أربعة مستويات تبدأ بالعامل العام General Factor (g) ، وتتدرج إلى العوامل الطائفية الرئيسة Major Group Factors التى تنقسم إلى عاملين أحدهما لغوى ـ تعليمى ، والآخر عملى ـ ميكانيكى ـ مكانى ، وينقسم كل من هذين العاملين بدوره إلى عوامل طائفية فرعية Minor Group Factors، وهذه تنقسم إلى عوامل نوعية Specific Factors تتعلق باختبارات معينة. ويُعد هذا التصنيف توليفًا من أعمال سبيرمان وثيرستون اللذين أشرنا إليهما من قبل. ويوضح الشكل التخطيطى (٩ ـ ٦) التالى هذا التصنيف الهرمى :



شكل (٩ ـ ٦) يوضح نموذج فيرنون للتصنيف الهرمى للقدرات العقلية

وقد اقترح كرونباك (Cronbach, 1970) بجامعة ستانفورد الأمريكية نموذجًا تصنيفيًا يعد تلخيصًا للنماذج السابقة وبخاصة نموذج فيرنون يوضحه الشكل التخطيطي (٩-٧) التالي:





شكل (٩ - ٧) يوضح بنية هرمية مقترحة للقدرات العقلية اقترحها كرونباك

ويتضح من شكل (٩ - ٧) أن العامل العام (g) أو الذكاء المائع الذى أشار إليه كاتل Cattell أو الذكاء التحليلى يقع فى قمة التصنيف، وهو مركز الشبكة العاملية بأسرها يليه العوامل العريضة المتعلقة به ، وتحتل العوامل النوعية قاعدة التصنيف.

الذكاء ومدخل تجهيز المعلومات ومعالجتها :

إن دراسة الذكاء لم تنفصل عن مبحث الفروق الفردية أكثر من نصف قرن من

الزمان، لذلك نلاحظ أن جميع النماذج العاملية لبنية العقل التى ناقشناها فيما سبق استندت استنادًا أساسيًا فى جمع البيانات وتحليلها إلى مدخل الفروق الفردية Differential Approach . ويتركز الاهتمام فى هذا المدخل على قياس الفروق بين الأفراد فى الذكاء والقدرات وليس على تفسير هذه الفروق.

وقد أدى ذلك إلى توجيه بحوث الذكاء فى الآونة الأخيرة وجهة جديدة تركز تركيزًا أساسيًا على الوظائف والعمليات المعرفية Cognitive Processes & Functions وهذا لا يعنى أن قياس الفروق الفردية ليس ضروريًا ، وإنما يعنى توجيه مزيد من الاهتمام إلى العمليات المعرفية التى تتحدد تجريبيًا.

وتستند الاختبارات العقلية إلى هذه العمليات بحيث يمكن أن تُزوِّدنا نتائجها بصفحات نفسية (بروفيل قدرات الأفراد) متعلقة بهذه العمليات تفيد فى التقويم التشخيصى والتعليم العلاجى للمهارات الأساسية. وبذلك توظَّف الاختبارات التشخيصية بدرجة أكبر من توظيف الاختبارات المعيارية التنبؤية. ولعلنا قد أوضحنا ذلك عند تمييزنا بين الاختبارات مرجعية المحك والاختبارات مرجعية الجماعة أو المعيار فى الفصل الثامن، عندما ذكرنا أن النوع الأول من الاختبارات يتميز بقيمة تشخيصية إذا حاولنا إحداث تكامل بين علم القياس وعلم النفس المعرفى فى بناء هذه الاختبارات، وهو ما سوف نوضحه فى الفصل الرابع عشر.

وقد أدى هذا التوجُّه الجديد إلى الاهتمام بمدخل تجهيز المعلومات ومعالجتها Information Processing Approach من أجل فهم أفضل لمجال الذكاء الإنسانى. ويسعى الباحثون فى هذا المجال إلى الكشف عن التمثيل Representations ، والعمليات Processes ، والاستراتيجيات Strategies التى يستخدمها الأفراد فى حل مشكلات محددة تحوز الاعتراف بأنها تتطلب «ذكاء» فى حلها.

كما يحاولون البحث عن كيفية إدماج كل من المدخل الفارق أو المدخل السيكومترى، ومدخل تجهيز المعلومات معًا فى دراسة الذكاء الإنسانى. وربما يرجع السبب فى ذلك إلى أن كلا من المدخلين يشوبه بعض القصور، فالمدخل الأول يغفل العمليات والاستراتيجيات التى تُشكل جزءًا كبيرًا بما نعنيه بالذكاء، أما المدخل الثانى فلا يزودنا بوسائل الدراسة المنظمة للمتغيرات المرتبطة بالفروق الفردية فى الأداء، وبخاصة أن هذا المدخل ربما يؤدى إلى تعظيم قيمة مكونات التجهيز Processing Components على الرغم من أن هذه المكونات تتعلق بأداء مهمة محددة - Task Specific .

فمن المعلوم أن الباحثين فى علم النفس الفارق يختلفون عن الباحثين فى علم النفس المعرفى فى استراتيجيات التنظير فى مجال الذكاء الإنسانى. فالباحثون فى علم النفس الفارق يحاولون التوصل إلى نظريات متكاملة للذكاء، أما الباحثون فى علم النفس المعرفى فإنهم يبدءون بنظريات فرعية صغيرة أو نماذج تتعلق بأداء مهام معينة، أو أحيانًا مجموعة من المهام.

ولعل مزايا أىٍّ من المدخلين تُعد عيوبا للآخر مما يتطلب مزيدًا من البحث حول كيفية إدماجهما والإفادة من ذلك فى تطوير بحوث الذكاء الإنسانى. ويرى سترنبرج (Sternberg, 1979) أن البحوث المعاصرة التى تعمل على التكامل بين المدخلين تتجه إلى بناء نظريات فرعية محدودة للذكاء Subtheories. وهذه النظريات تتخذ شكل نظريات جوانب من الذكاء، مثل القراءة، والاستدلال، والتصور البصرى المكانى، وما إلى ذلك، أو مجموعات من هذه الجوانب.

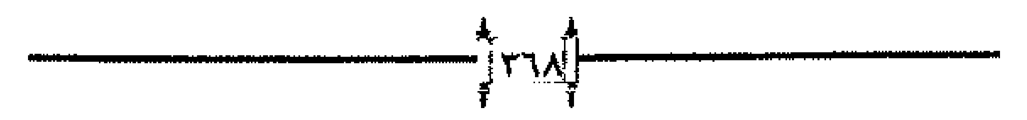
ولعل نظرية سترنبرج المتعلقة بتحليل مكونات الاستدلال القياسى Componential Theory of Analogical Reasoning وما تتضمنه من نماذج أساسية تتناول مظاهر القياس الاستدلالى Analogies والتصور الفكرى لعملية فحص البدائل Option Scanning تُعد مثالاً لهذا التوجه (Sternberg, 1977 , 1989) .

وهذه النظرية وأمثالها لا تفيد فقط فى فهمنا للذكاء ، وإنما فى تطوير اختبارات تستند إلى أفضل «المكونات Components» التى أطلق عليها سترنبرج «العمليات المعرفية الدقيقة Cognitive Microprocesses » التى يستخدمها الفرد فى حل مشكلات محددة يشتمل عليها اختبار أو ربما مفردة اختبارية.

وخلاصة ذلك أن تعدد نظريات ونماذج التكوين العقلى يُعد نتاجًا طبيعيًّا لتعقد وتشابك مفهوم الذكاء، ولا يجب أن تنظر إليها على أنها تناقش وجهات نظر متباينة، حيث إنها تعد نوافذ متعددة على ظاهرة الذكاء الإنسانى.

فهذه النماذج تزودنا بمنهجيات مختلفة لدراسة هذه الظاهرة التى تتطلب تكاتف جهود علماء النفس المعرفى وعلماء القياس النفسي والتربوى من أجل التوصل إلى نموذج جديد لبنية العقل استنادًا إلى نظرية معرفية .

ولعل هذا النموذج المأمول يسهم فى تعريف أفضل لما تقيسه اختبارات الذكاء بالفعل، ويقدم أساسًا أكثر رسوخًا لتقييم المضامين الاجتماعية والتربوية لهذه الاختبارت.



قياس الذكاء الإنساني :

بعد عرضنا السابق لأهم نظريات ونماذج التكوين العقلي وتوجهاتها المستقبلية ننتقل إلى كيفية قياس الذكاء الإنساني، فقد تبين لنا من هذا العرض مفهوم الذكاء العام وما يمكن أن يتضمنه من قدرات عقلية نستطيع قياسها بواسطة الاختبارات.

وسوف نتناول فيما يلي بعض اختبارات الذكاء العام الشائعة الاستخدام، ومعظم هذه الاختبارات تركز على بعض العوامل مثل الفهم اللغوى والاستدلال العام بدرجة أكبر من تركيزها على السرعة الإدراكية والتصور المكانى نظراً لأنه يُعتقد أن العوامل الأولى تمثل السلوك الذكى تمثيلاً أفضل.

لذلك يلاحظ أن محتوى كثير من اختبارات الذكاء العام يتعلق بعامل الفهم اللغوى، وعامل إجراء العمليات الحسابية، وأنواع مختلفة من عوامل الاستدلال، بالإضافة إلى محتويات متناثرة تتعلق بعوامل الذاكرة، والمكان، والإدراك.

وعلى الرغم من وجود ارتباطات مرتفعة بين اختبارات الذكاء العام تصل أحيانًا إلى ۸۰, ۰ بين بعض منها، إلا أن هذه الارتباطات ليست تامة مما يدل على أنها تقيس قدرات مختلفة إلى حد ما (Nunnally , 1978) . وهذا ربما يؤيد وجهات نظر بعض علماء النفس الذين يعارضون استخدام اختبارات الذكاء العام نظراً لأنه يمكن أن يتوصل الأفراد إلى حل مشكلة معينة باستخدام أساليب وعمليات مختلفة.

إذ ربما يعتمد أحد الأفراد اعتماداً أساسيًّا على قدرته المكانية فى حل تمارين هندسية، ويعتمد فرد آخر على حل هذه التمارين حلاً مجرداً اعتمادًا على قدرته الاستدلالية. لذلك ربما يقل استخدام اختبارات الذكاء العام مستقبلاً عندما تتوصل البحوث إلى طرق جديدة لبناء بطاريات من الاختبارات تقيس العوامل المتعددة للذكاء.

اختبارات الذكاء العام :

يمكن تصنيف اختبارات الذكاء العام إلى نوعين وفقًا لطريقة التطبيق. فالاختبارات التى تطبق على فرد واحد فى وقت واحد تسمى الاختبارات الفردية Individual Intelligence Tests،أما الاختبارات التى تطبق على عدد كبير من الأفراد فى وقت واحد فتسمى الاختبارات الجماعية Group Intelligence Tests ، ومعظم الاختبارات الفردية مثل مقياس ستانفورد بينيه Stanford - Binet ، ومقاييس ويكسلر Wechsler تحتاج إلى شخص مدرب على تطبيقها وتفسير نتائجها حيث يهتم بكيفية استجابة الفرد أو المفحوص. لذلك تطبق عادة الاختبارات الفردية بواسطة الأخصائيين

النفسيين بالمدارس والجامعات والمؤسسات حيث يُستفاد من نتائجها فى التشخيص الكلينيكى، واتخاذ قرارات مهمة مثل تشخيص حالات الضعف العقلى.

أما الاختبارات الجماعية ، فإنها تُعد أفضل وسائل تقييم القدرات العقلية لعدد كبير من المختبرين فى هذه المؤسسات بغرض التوجيه التعليمى والمهنى.

ونظراً لأنه يصعب بالطبع تناول جميع اختبارات الذكاء العام المستخدمة فى وقتنا الحاضر، فإننا سوف نوضح بإيجاز أمثلة لبعض الاختبارات الفردية والجماعية الشائعة الاستخدام.

أولاً: اختبارات الذكاء الفردية :

مقياس ستانفورد بينيه للذكاء(SBIS) Stanford - Binet Intelligence Scale :

سبق أن أشرنا إلى هذا المقياس فى سياق عرضنا لقياس الذكاء من منظور تاريخى. وهو يُعد من أشهر الاختبارات الفردية التى تقيس الذكاء العام وأكثرها استخداماً.

ويُعد مقياس ستانفورد ـ بينيه كما ذكرنا امتداداً لمقياس بينيه Binet الأصلى. وقد نشرت الطبعة الأولى لهذا المقياس عام ١٩١٦ فى الولايات المتحدة الأمريكية وأجريت عليه عدة تعديلات فى الأعوام ١٩٣٧، ١٩٦٠ ، ١٩٧٢، ١٩٨٤. وسوف نلقى الضوء فيما يلى على هذا المقياس:

مقياس ستانفورد ـ بينيه (طبعة عام ١٩١٦) :

قام تيرمان Terman بجامعة ستانفورد الأمريكية بترجمة مقياس بينيه ـ سيمون Binet - Semon وأجرى تطبيقه على عينة تشتمل على ١٠٠ طفل، ٤٠٠ فرد راشد. وكما فى مقياس بينيه الأصلى اعتمد تيرمان على ميزان العمر الزمنى Age Scale ، حيث قام بتجميع الاختبارات التى يشتمل عليها المقياس فى مستويات عمرية.

واشتمل المقياس على مفردات منتقاة من مقياس بينيه وإضافة مفردات جديدة تمثل عينة واسعة من المهام التى تتطلب ذكاء، ولكن هذه المهام لا تعتمد على خبرات مدرسية معينة، وأحد محكات انتقاء مفردة فى اختبار فرعى هو تزايد النسبة المئوية من الأطفال فى المستويات العمرية المتتالية الذين يستطيعون أن يجيبوا عنها إجابة صحيحة أى التمايز العمرى Age Differentiation،وتوضع المفردة عندئذ فى المستوى العمرى للاختبار الفرعى الأكثر ملاءمة له.

ومن المحكات الأخرى لانتقاء المفردات الارتباط بين درجات كل مفردة والدرجات الكلية في المقياس. وهذا يعني أن كل مفردة تقيس ما يتعلق بالمقياس ككل، وهو ما سبق أن أشرنا إليه باسم «تجانس المفردات Item Homogeneity» من حيث محتواها، وذلك في الفصل الرابع.

ويعتمد العمر العقلي Mental Age (MA)، ونسبة الذكاء Intelligent Quotient (IQ) في القياس على عدد الاختبارات الفرعية التي يجتازها المفحوص في المستويات العمرية المختلفة. وتحسب نسبة الذكاء باستخدام الصيغة :

$$\text{نسبة الذكاء (IQ)} = \frac{\text{العمر العقلي}}{\text{العمر الزمني}} \times ١٠٠ \qquad \ldots\ldots (١ـ٩)$$

وذلك بعد تحويل كل من العمر العقلي والعمر الزمني إلى شهور.

مقياس ستانفورد ـ بينيه (طبعة عام ١٩٣٧) :

أعيد تقنين المقياس على أطفال من مستويات عمرية مختلفة تبدأ من عام ونصف حتى خمسة أعوام ونصف بفاصل عمري قدره نصف عام بين كل مستوى وآخر، ومن أعمار تبدأ من ٦ أعوام حتى ١٤ عامًا بفاصل عمري قدره عام واحد، وكذلك من أعمار تبدأ من ١٥ عامًا حتى ١٨ عامًا بفاصل عمري قدره عام واحد. واشتملت العينة على الذكور والإناث من مستويات اجتماعية واقتصادية تفوق المتوسط، وبذلك لم تمثل المجتمع الأمريكي، وتم تجميع الاختبارات لكل فئة من هذه الفئات العمرية، وأضيف مستوى الراشد المتوسط، وثلاثة مستويات للراشد المتفوق.

ويوضح شكل (٩ ـ ٨) بعض المواد الاختبارية غير اللفظية التي يشتمل عليها مقياس ستانفورد ـ بينيه (Nunnally , 1970).



شكل (٩ ـ ٨) يوضح بعض المواد الاختبارية غير اللفظية في مقياس ستانفورد ـ بينيه

وعند اختبار طفل ينبغى على الفاحص أن يحدد المستوى العمرى القاعدى Basal Age (BA) وهو أعلى مستوى عمرى يستطيع عنده الطفل أن يجتاز جميع الاختبارات الفرعية المتعلقة به. وعندئذ يستمر اختبار الطفل حتى يصل إلى أعلى مستوى عمرى Ceiling Age (CA) يفشل عنده الطفل فى اجتياز أىٍّ من الاختبارات الفرعية المتعلقة به. ويحسب العمر العقلى بإضافة العمر القاعدى (BA) إلى عدد الشهور التى يحصل عليها الطفل من الاختبارات الفرعية التى اجتازها من أعلى مستوى عمرى وصل إليه (CA) ، حيث إن لكل من هذه الاختبارات عدد محدد من الشهور موضح فى دليل المقياس.

وبذلك يمكن حساب نسبة الذكاء (IQ) باستخدام الصيغة (٩ـ١) التى ذكرناها. ومن الجدير بالذكر أن المقياس أصبح يتكون من صيغتين متكافئتين (ل) ، (م) بعد أن قام تيرمان وميريل بإجراء التعديلات المناسبة عليه، وذلك إذا تطلب الأمر إعادة تطبيق المقياس على الفرد نفسه.

مقياس ستانفورد ـ بينيه (طبعة عام ١٩٦٠) :

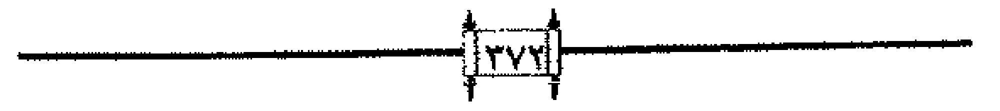
أعاد تيرمان مراجعة المقياس حيث انتقى أفضل مفردات كل من الصيغتين (ل) ، (م) واتسع المدى العمرى من عامين حتى الرشد.

واشتمل المقياس على مواد اختبارية تناسب الطفل الصغير مثل اللعب والصور وأشياء عيانية أخرى . وأعد دليل المقياس وكراسة تسجيل الإجابات وغير ذلك من المطبوعات التى يسترشد بها مستخدم المقياس فى تطبيقه، وفقًا للتعليمات المحددة بدقة.

ومن التعديلات الأساسية فى هذه الطبعة استخدام نسبة الذكاء الانحرافية Deviation IQ (DIQ) التى أشرنا إليها فى الفصل السادس، نظرًا لأوجه القصور المتعددة التى تتعلق بنسبة الذكاء (IQ) وبخاصة عدم تساوى الوحدات العمرية، وصعوبة تفسيرها فى المستوى العمرى للراشد، حيث إنه لا يوجد اتفاق على العمر الزمنى الذى يستخدم فى مقام صيغة نسبة الذكاء فى هذه الحالة، أى العمر الذى يُفترض أن يتوقف عنده النمو العقلى للفرد، حيث إن هذه تعد من المشكلات التى تحوطها اعتبارات كثيرة.

مقياس ستانفورد ـ بينيه (طبعة عام ١٩٧٢) :

لم تتضمن عينات الأفراد التى استندت إليها المعايير فى الطبعات السابقة فئات الأقليات Minority Group فى المجتمع الأمريكى، لذلك أعيد تقنين المقياس عام



١٩٧٢ على عينة مُمثلة لجميع الفئات بما فى ذلك الأقليات، واشتملت على ٢١٠٠ طفل بمعدل ١٠٠ طفل لكل مستوى عمرى (٢ - ١٨ عامًا) ، وظل متوسط نسبة الذكاء الانحرافية ١٠٠ وانحرافها المعيارى ١٦. ووُجد أن قيمة معامل ثبات درجات القياس باستخدام طريقة التجزئة النصفية تفوق ٠,٩٠ ، ولكن قيم معامل الارتباط بين درجات المقياس ، ودرجات تحصيل التلاميذ أو تقديرهم العام فى المواد الدراسية تراوحت بين ٠,٤٠ ، ٠,٧٥ ، وتعكس هذه القيم الموجبة حقيقة أن مقياس ستانفورد - بينيه مُشبع بمواد لفظية وبخاصة فى المستويات العمرية العليا، والدليل على ذلك أن قيم معامل الارتباط تزداد بالنسبة لمواد القراءة، واللغة الإنجليزية بينما تقل بالنسبة للرياضيات (Sax , 1964) .

مقياس ستانفورد - بينيه (طبعة عام ١٩٨٤) :

هذه الطبعة الرابعة للمقياس لا تختلف كثيرًا عن الطبعة السابقة فيما عدا بعض التغييرات المهمة حيث أُعدت للأغراض التالية :

(١) الإفادة فى التمييز بين الطلاب المتخلفين عقليًا والطلاب الذين يعانون من إعاقات معينة فى التعلم.

(٢) معاونة المربين والأخصائيين النفسيين فهم أسباب الصعوبات التى ربما تواجه بعض الطلاب فى المدرسة.

(٣) المعاونة فى التعرف على الطلاب الموهوبين.

(٤) دراسة نمو المهارات المعرفية لدى الأفراد من عمر عامين حتى الرشد.

وقد اشتمل المقياس على ١٥ اختبارًا فرعيًا تشكل نظامًا هرميًا يتكون من ثلاثة مستويات قمته تمثل الذكاء العام (g) يليه ثلاثة عوامل عريضة تمثل القدرات المتبلورة ، والقدرات المائعة - التحليلية، والذاكرة قصيرة الأمد، ويلى ذلك ثلاثة عوامل أقل اتساعًا هى الاستدلال اللفظى، والاستدلال الكمى، والاستدلال المجرد - البصرى، وهذه المستويات الهرمية الثلاثة موضحة بالشكل التخطيطى (٩ - ٩) التالى : (Thorndike , Hage & Sattler , 1986) :

العامل العام			
القدرات المتبلورة		القدرات المائعة ـ التحليلية	الذاكرة قصيرة الأمد
الاستدلال اللفظى	الاستدلال الكمى	الاستدلال المجرد ـ البصرى	
الكلمات	التعامل مع المقادير الكمية	تحليل الأنماط	ذاكرة الصور
الفهم	سلاسل الأعداد	مصفوفات	ذاكرة الجمل
السخافات	تكوين المعادلات	ثنى الأوراق وقطعها	ذاكرة الأرقام
العلاقات اللفظية			ذاكرة الأشياء

شكل (٩ ـ ٩) يوضح النظام الهرمى لمقياس ستانفورد ـ بينيه للذكاء (الطبعة الرابعة)

وقد اشتملت عينة التقنين على أكثر من ٥٠٠٠٠ فرد من مختلف الولايات تم تصنيفهم وفقًا للمنطقة الجغرافية، وحجم المجتمع المحلى، والسلالة، والنوع.

ووجد أن قيمة معامل استقرار الدرجات الكلية فى المقياس تفوق ٠,٩٠ وأكدت البيانات صدق التكوين الفرضى للمقياس ككل.

ويتميز المقياس بأنه يقيس نطاقًا واسعًا من عمليات تجهيز المعلومات لدى الفرد، ولكن لم تؤكد نتائج التحليل العاملى البنية الهرمية للتصنيف الموضح بشكل (٩ـ ٩).

لذلك ينبغى تفسير الدرجات الكلية فقط حيث إنها تدل على الذكاء العام. غير أن المقياس يتطلب مزيدًا من البحث لتحديد جوانب القوة والضعف عند تطبيقه على أطفال ما قبل المدرسة ، كما يتطلب أن تتضمن العينة أطفالاً من ذوى الاحتياجات الخاصة.

تعقيب على مقياس ستانفورد ـ بينيه :

على الرغم من أن مقياس ستانفورد ـ بينيه يُعد من أكثر مقاييس الذكاء العام

استخداماً، حيث أشار ثورنديك (Thorndike, 1975) إلى أنه طُبِّق على حوالى ۸۰۰,۰۰۰ فرد سنويًا ما بين عامى ١٩٦٠ وحتى ١٩٧٢ . ونال كثيرًا من الدراسات والبحوث حيث أشار الكتاب السنوى للمقياس العقلى عام ١٩٧٨ إلى ما يقرب من ١٥٩٠ مصدراً يتعلق بالمقياس (Buros, 1978) .

وبذلك يمكن القول بأنه حقق الغرض الذى بُنى من أجله، إلا أنه يُعاب عليه تركيزه الشديد على القدرة اللفظية . فتعليمات معظم أجزائه تقدم شفويًّا، وكثير من اختباراته الفرعية تتطلب استخدام الكلمات ، فإذا طبق المقياس على فرد لديه بعض الصعوبات اللغوية، فإنه ربما يصعب الحصول على صورة صادقة عن نموه العقلى، كما أنه يتطلب فاحصًا مدربًا تدريبًا عاليًا على استخدام المقياس لكى يتسنى له الحيطة من تدخل ظروف تؤدى إلى عدم دقة النتائج.

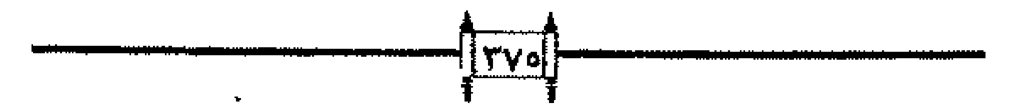
هذا بالإضافة إلى أن المقياس يعطى درجة كلية واحدة وبذلك لا يزودنا بمعلومات كافية عن القدرات المتعددة للفرد. وكذلك لا يناسب الكبار، حيث إن نوع الأداء العقلى الذى يمثله مفهوم العمر العقلى ينمو ببطء شديد بعد بلوغ السادسة عشر أو السابعة عشر من العمر مما يؤدى إلى مشكلات فى حساب نسبة الذكاء IQ بالنسبة للأعمار التى تزيد عن ذلك كما سبق أن أوضحنا.

ولعل استخدام نسبة الذكاء الانحرافية فى طبعة عام ١٩٦٠ وما بعدها هدف للتغلب على هذه المشكلة كوسيلة لجعل نسبة الذكاء (IQ) قابلة للمقارنة فى جميع المستويات العمرية.

ولعله يُعد من أفضل المقاييس السيكولوجية التى تتنبأ بدرجة جيدة بالتحصيل الدراسى. كما أن المقياس يُقدم للفاحص معلومات ثرية من خلال ملاحظته سلوك المفحوص أثناء حله للمشكلات وأدائه للمهام المتباينة فى درجة صعوبتها، كما يمكنه ملاحظة المدخل الذى ينتهجه المفحوص فى حل المشكلة مثل المحاولة والخطأ فى مقابل أسلوب التفكير المنتظم . لذلك يمكنه ملاحظة السلوك فى موقف مقنن إلى جانب الدرجة الكلية التى يحصل عليها المفحوص، ومقارنتها بالمعايير التى تم تحديثها.

مقاييس ويكسلر للذكاء Wechsler Intelligence Scales :

أعد ويكسلر Wechsler مجموعة من المقاييس الفردية للذكاء من منظور مختلف عن المنظور الذى استند إليه مقياس ستانفورد - بينيه ، حيث حاول تلافى أوجه قصور هذا المقياس التى أشرنا إليها فيما سبق.



ولعل أهم هذه الأوجه عدم ملاءمة مقياس ستانفورد - بينيه لقياس ذكاء الراشدين من حيث مفرداته ومعاييره، واقتصاره على درجة كلية واحدة للفرد الذى يطبق عليه الاختبار تدل على الذكاء العام، واعتماد معاييره فى الطبعات الأولى على مفهومى العمر العقلى ونسبة الذكاء اللذين كانا محل تساؤل ونقد مستمر وبخاصة فيما يتعلق بذكاء الراشدين، وكذلك تأكيده لعامل السرعة فى الإجابة عن مفرداته الموقوتة مما قد لا يناسب الراشدين.

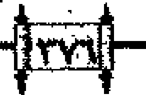
لذلك اهتم ويكسلر ببناء مقياس لذكاء الراشدين فى ضوء منظوره عن الذكاء كتجمعٍ من القدرات المترابطة وليس كقدرة عامة واحدة. ونظراً لأن مقياس ستانفورد - بينيه يعتمد اعتماداً أساسياً على الجانب اللفظى فيما عدا بعض المفردات فى المستويات العمرية المبكرة، فقد عمل على تضمين مقياسه جانبين أحدهما لفظى والآخر عملى. والهدف من ذلك التغلب إلى حد ما على مشكلة اللغة والثقافة والمؤثرات التعليمية (Wechsler , 1939) .

وبذلك يمكن قياس ذكاء الأفراد الذين لديهم صعوبات فى اللغة أو خبراتهم اللغوية محدودة، ويحصل الفرد فى المقياس على ثلاث درجات كلية تُمثل الأداء اللفظى، والأداء العملى، والذكاء العام. ففى هذا المقياس يمكن أن يحصل الفرد على نسبة الذكاء نفسها بطرق مختلفة بحسب المفردات التى يجتازها أو يُخفق فى الإجابة عنها.

وبذلك يمكن الإفادة من مقياس ويكسلر فى تشخيص أنماط تفكير الفرد بدرجة أفضل من مقياس ستانفورد - بينيه . كما يمكن الحصول على معلومات أكثر ثراء من خلال ملاحظة سلوك الفرد أثناء أدائه العملى. فإذا كان الأداء اللفظى يسمح للفاحص بملاحظة سلوك الفرد أثناء إجابته عن المفردات اللفظية، فإن الأداء العملى بما يتطلبه من تركيز وانتباه ووقت أطول يُمكِّن الفاحص من ملاحظة توجُّه الفرد ومدخله فى حل المشكلات. لذلك فإن مقياس ويكسلر يستخدم كأداة مناسبة فى التشخيص الكلينيكى.

وقد صدر مقياس ويكسلر للراشدين أول مرة عام ١٩٣٩، واشتمل على صورتين متكافئتين، وأُطلق عليه «مقياس ويكسلر - بلفيو للذكاء» نسبة إلى مستشفى بلفيو للأمراض النفسية كما ذكرنا من قبل.

ولكن نظراً لأن الجماعة المعيارية التى استخدمت فى تقنين الاختبار وتحديد معاييره كانت غير ممثلة للمجتمع الأمريكى ، فقد أعاد ويكسلر تقنينه مرة أخرى عام ١٩٥٥ على عينة أكثر تمثيلاً ، وأطلق عليه « مقياس ويكسلر لذكاء الراشدين



Wechsler Adult Intelligence Scale (WAIS)،ثم أعيد تقنينه وحُـددت له معايير جديدة عام ١٩٨١ (Wechsler , 1981) ، وأطلق عليه (WAIS - R) .

وقـد قام ويكـسـلر عـام ١٩٤٩ ببناء مقياس آخـر لذكاء الأطـفال، Wechsler Intelligence Scale for Children (WAIS) ، وذلك للمدى العمرى ٥ ـ١٥ عامًا، كما قام عام ١٩٦٧ ببناء مقياس ثالث لذكاء أطفال ما قبل المدرسة

Wechsler Preschool and Primary Scale of Intelligence (WPPSI)

وذلك للمدى العمرى ٤ ـ ٦ أعوام. وسوف نوضح فيما يلى بإيجاز كل من هذه المقاييس:

مقياس ويكسلر لذكاء الراشدين (WAIS - R) :

صمم هـذا المقياس لـقياس ذكاء الـراشدين بخاصـة وذلك فى المدى العـمرى ١٦ـ٧٥ عامًا،لذلك راعى ويكسلر أن يـكون محتوى المقياس وطريقة تصحـيحه مناسبة لهم، وأن تكون معاييره مستمده من عينات من الراشدين. وقد تبنّى تعريفًا للذكاء يؤكد القدرة العامة كما أشرنا فى مستـهل هذا الفصل، حيث أوضح أن الذكاء هو قدرة الفرد الكلية على العمل الهادف، والتفكير المنطقى، والتعامل مع بيئته بفاعلية، ولذلك اعتبر أن الاختبارات الفرعية التى اشتمل عليها المقياس لا تقيس أنواعًا مختلفة من الذكاء.

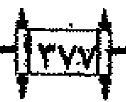
ومع هذا فإنه أكد الإفادة من المقياس كـأداة تشخيص وذلك باستخدام أنماط أو بروفيل درجات الاختبارات الفرعية التى يشتـمل عليها المقياس كأساس للاستدلال على إمكانات الفرد العقلية وحالته الانفعالية.

ولكنه أشار إلى أن الذكاء لا يُعدُّ مجموع الـقدرات التى تكشف عنها الاختبارات الفرعية المختـلفة، وإنما يعدُّ تكوين متناسق من القدرات إضافة إلـى العوامل الانفعالية والشخصية التى تؤدى إلى السلوك الذكى (Wechsler, 1958).

ويشتمل المقياس على مقياسين فرعيين أحدهما المقياس اللفظى Verbal Scale والآخر مقياس الأداء العملى Performance Scale .ويتكون المقـياس الأول من ستة اختبـارات فرعية، والـمقياس الثـانى من خمسـة اختبارات فـرعية. وهذه الاختـبارات كالتالى:

المقياس اللفظى:

(١) المعلومات : أسئلة تتنـاول معلومات عامة فى مـوضوعات مختلفـة مرتبة



تصاعديًا من حيث درجة صعوبتها، وهى مصممة لقياس المعلومات العامة المتوافرة لدى معظم قطاعات المجتمع.

(٢) **الفهم** : أسئلة تقيس الحكم الواقعى والمعلومات.

(٣) **الحساب** : مسائل كلامية تتطلب قدرًا ضئيلاً من المهارات الحسابية، وتقيس التركيز.

(٤) **المتشابهات** : أسئلة تتضمن أزواجًا من الكلمات، وتتطلب توضيح أساس التشابه بين كل زوج منها، وتقيس التفكير المجرد.

(٥) **وسع الذاكرة للأرقام** : أسئلة تتطلب إعادة ذكر أرقام بترتيبها الأصلى أو بالترتيب العكسى.

(٦) **معانى الكلمات** : أسئلة تتطلب تعريف كلمات أو مصطلحات عامة، وتقيس مستوى فهم الكلمات.

مقياس الأداء العملى:

(١) **رموز الأرقام** : مهام تتطلب استبدال شىء برمز مجرد استرشادًا بقائمة توضح التناظر بينهما، وذلك فى مدة زمنية تتراوح بين دقيقة ودقيقة ونصف، وتقيس الوظائف البصرية والحركية.

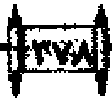
(٢) **تكميل الصور**: تتطلب تحديد الجزء المفتقد من صورة معينة، ويقيس إدراك التفاصيل.

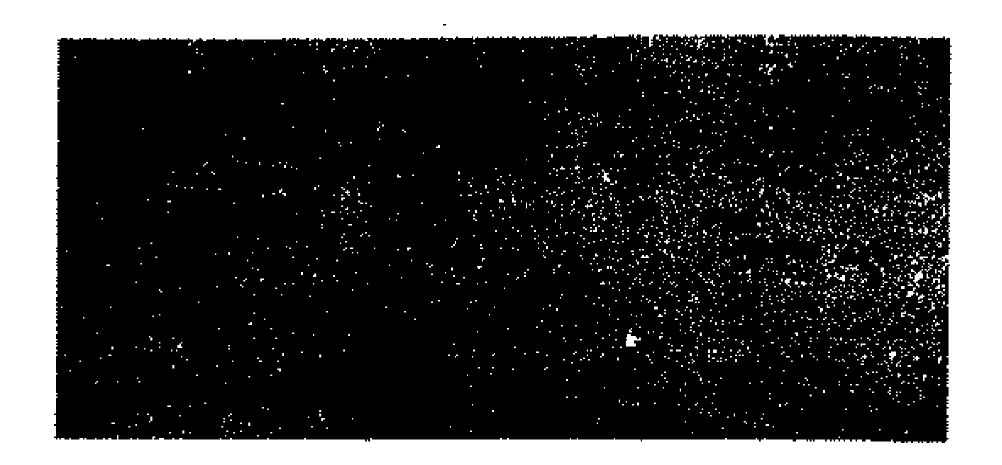
(٣) **تصميم المكعبات** : يتطلب ترتيب مكعبات ذات ألوان معينة بحيث تكون مماثلة لتصميمات معطاة، وتقيس القدرة على التخطيط.

(٤) **ترتيب الصور** : تتطلب ترتيب بطاقات مختلفة ترتيبًا متسلسلاً ، وتقيس الاستدلال غير اللفظى.

(٥) **تجميع الأشياء** : يتطلب تجميع أجزاء متاهات متباينة فى درجة تعقدها بأسرع ما يمكن، ويقيس القدرة على تحليل علاقة الأجزاء بالكل.

ويوضح شكل (٩ ـ ١٠) بعض المواد الاختبارية التى يشتمل عليها مقياس ويكسلر للراشدين (Horrocks , 1964).





شكل (٩ ـ ١٠) يوضح بعض المواد الاختبارية فى مقياس ويكسلر للراشدين

وقد تم ترتيب مفردات كل من هذه الاختبارات الفرعية ترتيبًا تصاعديًا بحسب مستوى صعوبتها، ويجرى تطبيقها وفقًا لهذا الترتيب قدر الإمكان، وينتهى التطبيق إذا أخفق الفرد عددًا محددًا من المرات المتتالية. وبعض هذه المفردات ينبغى الإجابة عنها فى زمن محدد، بينما البعض الآخر غير موقوت.

وينبغى على الفاحص الذى يستخدم المقياس أن يكون على درجة عالية من المهارة والتدرب، وأن يراعى اتباع التعليمات التفصيلية الموضحة فى دليل المقياس بدقة. ويستغرق تطبيق المقياس ساعة تقريبًا، وهو زمن أقل مما يتطلبه تطبيق اختبار ستانفورد ـ بينيه.

ويتم تصحيح الاختبارات الفرعية بمقارنتها بقائمة الاستجابات المقبولة المحددة فى دليل المقياس. وتقدَّر الدرجات أو النقاط على أساس درجة صحة الإجابة، وتُراعى أيضًا سرعة الاستجابة فى بعض الاختبارات الفرعية.

لذلك يعتمد مقياس ويكسلر على النقاط Point Scale وليس على العمر Age Scale كما فى مقياس ستانفورد ـ بينيه.

وتحوَّل الدرجات الخام الكلية فى كل اختبار فرعى إلى درجات معيارية اعتدالية متوسطها ١٠ وانحرافها المعيارى ٣، وتجمع الدرجات الكلية للاختبارات الفرعية، ثم تحوَّل الدرجات المعيارية الاعتدالية لكل من المقياس اللفظى، ومقياس الأداء العملى،

والمقياس ككل إلى معيار «نسبة الذكاء الانحرافية DIQ» التى أوضحناها فى الفصل السادس (شكل ٦ ـ ١) وقد حدد ويكسلر متوسط هذه النسبة ١٠٠ وانحرافها المعيارى ١٥ نقطة.

وهذا يعنى أن الفرد الذي تنحرف درجته الخام فى أحد المقياسين أو المقياس ككل وحدة انحراف معيارى واحدة أعلى متوسط مجموعته العمرية تكون نسبة ذكائه ١١٥ مثل فرد فى أى مجموعة عمرية أخرى تنحرف درجته الخام أيضًا وحدة انحراف معيارى أعلى متوسط مجموعته. واستُمدت معايير المقياس من عينة مستعرضة تُمثل مجتمع الراشدين بالولايات المتحدة الأمريكية. وبإجراء هذا التحويل يمكن المقارنة بين الدرجات الكلية لكل من الاختبارات الفرعية على ميزان موحّد سواء للفرد نفسه أو بين الأفراد.

ومما يُميز نسبة الذكاء الانحرافية أنه يمكن حسابها لكل مجموعة عمرية على حدة. وهذا بالطبع يحدث فى مقياس ذكاء الأطفال، ولكن يعد أمرًا جديدًا فى حالة الراشدين، فمثلاً تبين أن المجموعة العمرية ٢٥ ـ ٢٩ وفقًا لهذا المعيار تؤدى أداء أفضل من أى مجموعة عمرية أخرى.

لذلك فإن أفراد هذه الأعمار ينبغى أن يحصلوا على درجات خام فى المقياس أعلى من أفراد الأعمار الأخرى لكى يحصلوا على نسبة الذكاء الانحرافية نفسها.

وقد استُمدت معايير الطبعة الأخيرة للمقياس من عينة مستعرضة قوامها ١٨٨٠ فردًا موزعة على تسع مجموعات عمرية تمّ اختبارهم فى المدة بين عامى ١٩٧٧ ، ١٩٨٠.

وفيما يتعلق بالخصائص السيكومترية للمقياس بيَّنت الدراسات أن الدرجات الكلية للمقياس تتميز بقدر مرتفع من الثبات، غير أن الدرجات الكلية للاختبارات الفرعية أقل ثباتًا نظرًا لقلة عدد مفرداتها، وقيم معامل الاتساق الداخلى للمقياس اللفظى أعلى من قيم هذا المعامل لمقياس الأداء العملى. وعلى الرغم من أن استخدام درجات المقياسين كل على حدة يتطلب أن تكون قيم معامل الارتباط بينهما منخفضة ، إلا أن ويكسلر وجد أن هذه القيم مرتفعة وتصل أحيانًا إلى ٠,٨١ للأفراد فى المجموعة العمرية ٤٥ – ٥٤ عامًا مما ينفى افتراض أن كلا من المقياسين مستقل عن الآخر.

ولهذه الأسباب أوضح ويكسلر أن تفسير بروفيل درجات الأفراد الأسوياء فى الاختبارات الفرعية يكون غير ذى معنى. فأى فرق بين درجات الفرد فى أى اختبارين فرعيين ينبغى أن يكون كبيرًا قبل أن نعتبره فرقا دالا، ومع هذا فإن الفروق الكبيرة بين

الدرجات المحوّلة للاختبارات الفرعية، وكذلك بين نسب الذكاء الانحرافية للمقياس اللفظي ومقياس الأداء العملي يمكن الإفادة منها في تشخيص بعض الاضطرابات العقلية.

أما فيما يتعلق بصدق المقياس، فلا توجد أدلة كافية بهذا الشأن، ولكن يوجد ارتباط مرتفع بين درجات مقياس ويكسلر ومقياس ستانفورد - بينيه يبلغ حوالى ٠,٨٥ للمقياس ككل، ٠,٨٠ للمقياس اللفظى، ٠,٦٩ لمقياس الأداء العملى (Wechsler , 1958).

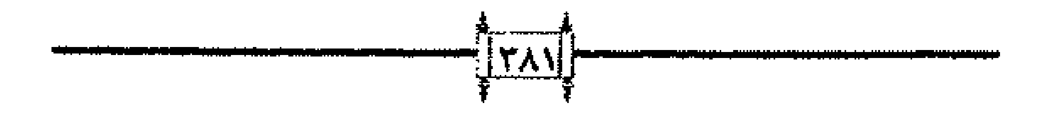
وقد أشار ويكسلر إلى صدق محتوى المقياس من حيث الوظائف التى تقيسها الاختبارات الفرعية وملاءمتها لتعريف الذكاء الذى تبنّاه. وبيّنت دراسات صدق التكوين الفرضى فيما يتعلق بالارتباطات بين المقاييس الفرعية، وبين كل من المقياس اللفظى ومقياس الأداء العملى أن مقياس ويكسلر يقيس الذكاء العام، وأكد ذلك دراسات التحليل العاملى. غير أن التقييم الدقيق للخصائص السيكومترية للمقياس تتطلب مزيداً من البحث.

مقياس ويكسلر لذكاء الأطفال (WISC):

يعد هذا المقياس امتداداً للصيغة الأولى لمقياس ويكسلر - بلفيو لذكاء الراشدين، بحيث يقيس ذكاء الأطفال فى المدى العمرى ٥ - ١٥ عامًا، ١١ شهرًا، والطبعة الأولى لهذا المقياس تم إصدارها عام ١٩٤٩. واشتمل المقياس على خمسة اختبارات فرعية تقيس الأداء اللفظى، وخمسة اختبارات أخرى تقيس الأداء العملى وإضافة اختبارين تكميليين لتطبيقهما إذا تبين صعوبة أحد الاختبارات الفرعية الأخرى بالنسبة للطفل المُختبر.

والاختبارات الفرعية التى يشتمل عليها مقياس الأداء اللفظى هى : المعلومات العامة، والفهم العام، والحساب، والمتشابهات، ومعانى الكلمات، واختبار بديل يقيس وسع ذاكرة الأرقام. أما الاختبارات الفرعية التى يشتمل عليها مقياس الأداء العملى فهى: تكميل الصور، وترتيب الصور، وتصميم المكعبات، وتجميع الأشياء، والترميز، واختبار بديل يشتمل على متاهات.

وهذه الاختبارات الفرعية تُناظر اختبارات الراشدين من حيث المحتوى والوظائف العقلية التى تقيسها، ومفردات هذه الاختبارات مرتبة بحسب مستوى صعوبتها، ومجمعة معًا وفقًا لمحتواها. ويُوقف تطبيق أى اختبار فرعى إذا أخطأ الطفل فى إجابته عدة مرات متتالية، ولكن هذا العدد أقل مما هو فى حالة مقياس الراشدين، وذلك



لتقليل شعور الطفل الصغير أو الأقل ذكاء بالإحباط أو الفشل، وتقليل زمن التطبيق الذى يستغرق ساعة واحدة تقريباً.

وقد أُجريت بعض التعديلات على المقياس وصدرت طبعته الثانية عام ١٩٧٤ وأطلق عليه (WISC - R) ، ولا يختلف كثيراً عن طبعته الأولى فيما عدا تغيير ترتيب تطبيق الاختبارات الفرعية، وتوسيع مدى الأعمار عامًا واحدًا ليشمل العمر (١٦ عاما، ١١ شهراً ، ٣٠ يومًا) ، وأدى ذلك إلى بعض التداخل مع مقياس الراشدين الذى يبدأ بالمستوى العمرى ١٦ عامًا حيث ينتهى مقياس الأطفال، وأُعد دليل خاص بتفسير درجات المقياس.

وقد أعيد تقنين الاختبار فى هذه الطبعة بحيث أصبحت العينة أكثر تمثيلاً لمجتمع الأطفال فى الولايات المتحدة الأمريكية، واشتملت على ٢٢٠٠ طفل من مختلف المناطق الجغرافية، والمستويات الاجتماعية والاقتصادية، والعمر، والنوع.

وقد صدرت الطبعة المعدلة الثالثة للمقياس عام ١٩٩١ (Wechsler , 1991) تحت مسمى (WISC III) التى تعد أكثر الطبعات استخدامًا لقياس ذكاء الأطفال، وأجرى حولها كثيراً من البحوث فى مجال التربية الخاصة. وقد أُضيف فى هذه الطبعة اختبار تكميلى فى مقياس الأداء العملى هو اختبار البحث عن الرموز، حيث يُطلب من الطفل الفحص البصرى لمجموعة من الرموز ويُحدد وجود رمز معين، وبذلك أصبح عدد الاختبارات الفرعية عشرة اختبارات، وثلاثة اختبارات تكميلية.

كما أن الاختبارات الفرعية التى تقيس الأداء العملى موقوتة، وقد تم تحويل درجات كل اختبار فرعى إلى درجات معيارية متوسطها ١٠ وانحرافها المعيارى ٣ كما فى اختبارات مقياس الراشدين، وذلك لكى يتسنى المقارنة المباشرة بين درجات الاختبارات الفرعية من خلال تحليل الصفحة النفسية (بروفيل) لدرجات الطفل على الرغم من النقد الذى وجِّه مؤخراً إلى هذه الصفحات النفسية Profiles. ويمكن الحصول على نسب الذكاء الانحرافية لكل من الأداء اللفظى والأداء العملى والذكاء العام التى متوسطها ١٠٠، وانحرافها المعيارى ١٥، وأُعدت جداول خاصة يستخدمها الفاحص لتفسير انتشار درجات الاختبارات الفرعية، والفروق فى الأداء اللفظى، ودلالة الفروق بين الاختبارات الفرعية التى تختلف عن المتوسط العام لدرجات جميع هذه الاختبارات. كذلك تضمنت هذه الجداول مؤشرات لمكافئ الدرجات Score Equivalents وذلك لأربعة عوامل جديدة تم التوصل إليها عن طريق التحليل العاملى (Kaufman , 1979). وهذه العوامل هى :

(١) الفهم اللغوى : ويتضمن المعلومات، والمتشابهات، ومعانى الكلمات، والفهم العام.

(٢) التنظيم الإدراكى : ويتضمن تكميل الصور، وترتيب الصور، وتصميم المكعبات ، وتجميع الأشياء.

(٣) التحرر من مؤثرات التشتت : ويتضمن الحساب ، ووسع ذاكرة الأرقام.

(٤) سرعة تجهيز المعلومات : ويتضمن الترميز، والبحث عن الرموز.

وقد أعدت معايير الاختبار استناداً إلى عينة مماثلة للعينة التى استخدمت فى الطبعة الثانية للمقياس، وروعى تساوى عدد كل من الذكور والإناث.

وأوضحت نتائج تقدير الثبات أن قيم معامل الاستقرار لجميع الاختبارات الفرعية فيما عدا الاختبارين التكميليين فى مقياس الأداء العملى تراوحت بين ٦٠,٠ ، ٩٥,٠ ولنسب الذكاء الانحرافية المناظرة لدرجات كل من المقياسين الفرعيين والمقياس ككل حوالى ٩٠,٠ ، كما تراوحت قيم معامل الاستقرار للعوامل الأربعة التى أشرنا إليها بين ٨٠,٠ ، ٩٥,٠ .

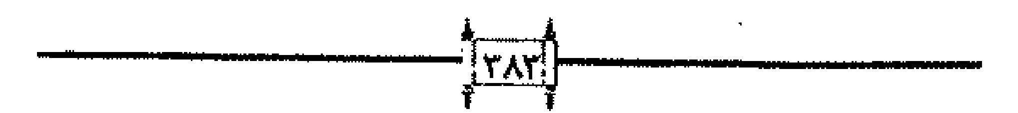
أما فيما يتعلق بصدق المقياس، فقد قدمت نتائج التحليل العاملى بعض الأدلة عن صدق التكوين الفرضى للمقياس، ولكن لا توجد مؤشرات عن الصدق المرتبط بمحك، حيث إن الدراسات لا تزال جارية فى هذا الشأن.

وينبغى عند تفسير درجات الطفل فى الاختبارات الفرعية مراعاة المتغيرات التى يمكن أن تؤثر فى هذه الدرجات، فدرجات الطفل فى الاختبار الفرعى المتعلق بالحساب ربما تتأثر ببعض المتغيرات مثل المهارة الحسابية للطفل، وقدرته على الاستدلال العددى، ومدى تشتت انتباهه، والمعلومات المكتسبة من المدرسة، ومستوى القلق لديه، والقدرة على العمل فى حدود زمن معين وغير ذلك.

مقياس ويكسلر لذكاء أطفال قبل المدرسة (WPPSI):

أعد ويكسلر أيضاً هذا المقياس عام ١٩٦٧، بهدف قياس ذكاء الأطفال الذين تتراوح أعمارهم بين أربعة أعوام وستة أعوام ونصف، ولا يختلف هذا المقياس كثيراً عن مقياس ذكاء الأطفال (WAIS - III) فيما عدا إضافة ثلاثة اختبارات فرعية تناسب طفل ما قبل المدرسة هى:

(١) بيت الحيوان : وهو اختبار موقوت يتطلب من الطفل وضع أسطوانة ملونة فى فتحة مناسبة أمام الحيوان.



(٢) تصميم هندسى : وهو اختبار يتطلب مهارة إدراكية حركية.

(٣) الجمل : وهو اختبار تكميلى لقياس التذكر الفورى.

كذلك لا يختلف هذا المقياس عن مقياس ذكاء الأطفال سواء فى طريقة بنائه أو تطبيقه أو تصحيحه.

وقد استندت معايير هذا المقياس على عينة قومية ممثلة تمثيلاً جيدًا للمناطق الجغرافية ومهنة الآباء، والنوع،والسلالة، واشتملت على ٢٠٠ طفل من الذكور والإناث بالتساوى فى كل من المستويات العمرية الستة التى يفصل بين كل منها نصف عام. ونحصل عن طريق المقياس على نسبة الذكاء الانحرافية التى متوسطها ١٠٠ وانحرافها المعيارى ١٥ لكل من المقياس اللفظى، ومقياس الأداء العملى، والمقياس ككل.

وقد بينّت الدراسات أن قيم معامل الاتساق الداخلى لكل من المقياسين الفرعيين والمقياس ككل كانت فى حدود ٠,٩٠ أو أقل ، مما يدل على ارتفاع قيمة ثبات درجة كل منها، كما تراوحت قيم معامل الاستقرار بفاصل زمنى ثلاثة أشهر لعمر خمسة أعوام بين ٠,٨٠ ، ٠,٩٠ . كذلك وُجد أن قيمة معامل الارتباط بين الدرجات الكلية فى المقياس والدرجات الكلية فى مقياس ستانفورد - بينيه (طبعة عام ١٩٦٠) بلغت ٠,٧٦ مما يدل على أن المقياسين يتفقان إلى حد ما فيما يقيسانه.

مقارنة بين مقاييس ويكسلر ومقياس ستانفورد - بينيه :

إن كثيرًا من ميزات مقياس ستانفورد - بينيه تتوافر فى مقاييس ويكسلر ، وكما أوضحنا فإن المقياسين مرتبطان ارتباطًا كبيرًا فى معظم المستويات العمرية. وعلى الرغم من أن كلا من المقياسين ربما يستخدم بدرجة أفضل فى أغراض معينة، إلا أن اختيار أى منهما يعتمد على ما يُفضله مستخدم المقياس، لذلك ربما يكون من المفيد أن توضح فيما يلى بعض الفروق الأساسية بين كل من المقياسين ليسترشد بها فى اختياره:

(١) إن الأساس المنطقى الذى بُنى عليه مقياس ويكسلر هو الحصول على قياس للذكاء العام للراشدين بخاصة، ومن ثم اعتمد على هذا المقياس فى بناء مقياسيه الآخرين لذكاء الأطفال، فالمقياسان مماثلان لمقياس الراشدين فيما عدا بعض الاختبارات الفرعية التى تناسب الأطفال، وأحدهما مصمم للأطفال الذين تتراوح أعمارهم بين ٥ ، ١٧ عامًا ، والآخر لأطفال ما قبل المدرسة الذين تتراوح أعمارهم بين ٤ ، ٦,٥ عامًا.

أما مقياس ستانفورد ـ بينيه فهو مصمم أساسًا لقياس الذكاء العام للأطفال الذين تتراوح أعمارهم بين ٢، ١٨ عامًا ، ويمكن تطبيقه على الراشدين. لذلك فإنه يُفضل عن مقياس ويكسلر في قياس ذكاء أطفال ما قبل المدرسة. وجميع هذه المقاييس تُعد من المقاييس الفردية.

(٢) مقاييس ويكسلر تشتمل على اختبارات فرعية مرتبة وتطبق وفقًا لهذا الترتيب، بينما الاختبارات التي يشتمل عليها مقياس ستانفورد ـ بينيه مرتبة وفقًا للمستويات العمرية.

(٣) مقاييس ويكسلر تتضمن مهامًا لفظية وأخرى عملية، بينما محتوى مقياس ستانفورد ـ بينيه مثقل بالجانب اللفظي.

(٤) مقاييس ويكسلر تزودنا بثلاث نسب للذكاء تتعلق بالجانب اللفظي، والأداء العملي، والمقياس ككل، وكذلك عشر درجات للاختبارات الفرعية، بينما يقتصر مقياس ستانفورد ـ بينيه على نسبة ذكاء كلية واحدة إضافة إلى درجة تمثل العمر العقلي.

(٥) تُعد مقاييس ويكسلر من نوع الموازين التي تعتمد على النقاط Point Scale، أما مقياس ستانفورد ـ بينيه فيُعد من الموازين التي تعتمد على العمر Age Scale.

(٦) يمكن استخدام مقاييس ويكسلر كأدوات تشخيصية في المواقف المدرسية والكلينيكية، غير أن هذا يتطلب فاحصًا مدربًا تدريبًا عاليًا على كيفية استخدام هذه المقاييس وتفسير درجاتها.

(٧) على الرغم من عدم وجود أدلة كافية عن مدى تنبؤ مقاييس ويكسلر بالتحصيل الدراسي والنجاح المهني، إلا أن الأدلة المتوافرة أوضحت أن هذه المقاييس لا تختلف في ذلك عن مقياس ستانفورد ـ بينيه.

(٨) يصعب باستخدام مقاييس ويكسلر قياس المستويات الدنيا للضعف العقلي وتؤدي إلى نسب ذكاء أكبر من حقيقتها عند هذه المستويات. ومع هذا فإن سقف المقاييس منخفض نسبيًا في المستويات العليا، وتؤدي إلى نسب ذكاء أقل من حقيقتها في هذه الحالات. وعلى الرغم من اشتمال هذه المقاييس على مهارات لفظية وعملية، إلا أنها مثل مقياس ستانفورد ـ بينيه تقيس مجموعة محدودة من العوامل أو القدرات التي تُمثل السلوك الذكي.

ومن الجدير بالذكر أن مقياس ويكسلر ـ بلفيو لذكاء الراشدين والمراهقين ومقياسه لذكاء الأطفال، وكذلك مقياس ستانفورد ـ بينيه أجرى تعريبها وتقنينها في البيئة المصرية.

بعض مقاييس الذكاء الفردية الأخرى:

توجد مقاييس فردية متعددة أخرى لقياس الذكاء، ولكنها ربما تكون أقل شيوعًا من مقاييس ويكسلر، ومقياس ستانفورد - بينيه. وبعض هذه المقاييس تطبق على الأسوياء، والبعض الآخر يناسب الأفراد من ذوى الإعاقات الحسية أو الجسمية أو اللغوية، فقد لاحظنا أن عينات تقنين المقاييس السابقة لم تتضمن أمثال هؤلاء الأفراد.

وعلى الرغم من أن معظم مقاييس الذكاء الفردية الأخرى لا ترقى فى خصائصها السيكومترية، ونوعية عينات تقنينها وكفايتها، واكتمال أدلتها وتعليماتها، إلى مستوى مقاييس ويكسلر وستانفورد - بينيه، إلا أنهما تفيد فى أغراض معينة، أو لفئات خاصة. وكثير منها لا يعتمد بدرجة كبيرة على الاستجابات اللفظية التى اهتمت بها المقاييس السابقة، وإنما تتطلب أن يشير الفرد ، أو يستجيب بـ (نعم) أو (لا) دون الحاجة إلى إحداث تكامل بين الوظائف البصرية والحركية التى تتطلبها الاختبارات الفرعية لمقاييس ويكسلر.

كذلك يشتمل كثير منها على اختبارات لقياس الأداء العملى، ولعل النقص الملحوظ فى هذا الجانب العملى فى مقياس ستانفورد - بينيه هو الذى شجع على بناء هذه المجموعات الأخرى من مقاييس الذكاء الفردية.

ونظرًا لأن بعض هذه المقاييس صُمّمت لأغراض معينة أو لفئات خاصة، فإن استخداماتها تكون مقصورة على هذه الفئات، وهذا يحد من تنوع الوظائف والإمكانات العقلية التى يمكن أن تقيسها، ومع هذا فإنها تعد أكثر ملاءمة لهذه الفئات من المقاييس الرئيسة السابقة.

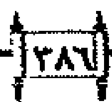
غير أنه فى هذه الحالة لا نستطيع مقارنة نسب الذكاء التى نحصل عليها من هذه المقاييس الأخرى بنسب الذكاء التى نحصل عليها من المقاييس الرئيسة. ويمكن اعتبار نتائج هذه المقاييس الأخرى مكملة لنتائج تطبيق مقاييس ويكسلر وستانفورد - بينيه، حيث تفيد فى أغراض الفحص والمتابعة وإعادة التقييم، أو إذا لم يكن هناك وقت متاح لتطبيق أحد هذين المقياسين الرئيسيين .

وفيما يلى أسماء بعض هذه المقاييس :

(١) بطارية التقييم للأطفال التى أعدها كوفمان عام ١٩٨٣.

Kaufman Assessment Battery for Children

وهى مجموعة اختبارات فردية لقياس كل من الذكاء والتحصيل وتناسب الأطفال الذين تتراوح أعمارهم بين ٢,٥ ، ١٢,٥ عامًا ،وتتميز بتركيز مفرداتها على العمليات



العقلية التى يستخدمها الطفل فى حل المشكلات بدلاً من تركيزها على المحتوى، ولا يعتمد قياس هذه العمليات على ما تعلمه الطفل فى المواد الدراسية. وقد استند بناء الاختبارات الفرعية إلى النظريات السيكولوجية ونظريات سيكولوجية الأعصاب. وتتضمن الاختبارات مفردات لفظية وغير لفظية، وأعدت معايير منفصلة للأطفال الأسوياء والأطفال من ذوى الاحتياجات الخاصة فى البيئة الأمريكية.

وتشتمل البطارية على ١٦ اختباراً فرعيًا، عشرة منها تتعلق بالذكاء العام، والستة الباقية تتعلق بالتحصيل.

وتنقسم اختبارات الذكاء العام العشرة إلى مجموعتين تتطلب إحداهما اختبارات التجهيز المتتابع للمعلومات Sequential Processing، والأخرى تتطلب التجهيز الآنى للمعلومات Simultaneous Processing. وكلاهما يتطلب من الطفل توليف أو إحداث تكامل بين معلومات مكانية أو متشابهات.

أما الاختبارات الفرعية التى تتعلق بالتحصيل فتشتمل على مفردات تتضمن كلمات تعبيرية، وحساب، وألغاز، وقراءة، وفهم المقروء.

ولا تطبق جميع الاختبارات الفرعية على الأطفال فى المستويات العمرية المختلفة، وإنما تطبق سبعة اختبارات فرعية على الأطفال الذين تبلغ أعمارهم ٢,٥ عامًا، وثلاثة عشر اختباراً على الأطفال فى المستويات العمرية ٧ ـ ١٢,٥ عامًا، ويمكن الحصول على درجتين إحداهما تشير إلى التجهيز المتتابع والأخرى تشير إلى التجهيز الآنى للمعلومات.

ونظراً لحداثة هذه البطارية ، فإنها تتطلب مزيداً من البحث، فيما يتعلق بصدق قرارات التشخيص والتسكين فى الإطار المدرسى استناداً إلى نتائجها.

(٢) بطارية الاختبارات السيكولوجية التربوية التى أعدها وودكوك، وجونسون عام ١٩٧٧، وأجرى تعديلها فى طبعتها الثانية عام ١٩٨٩.

Woodcock - Johnson Psychoeducational Battery (WJ - R)

وتشتمل على مجموعة متسعة من الاختبارات التى تطبق فرديًا على الأفراد الذين تتراوح أعمارهم بين عامين ، ٩٥ عامًا، وتقيس الاختبارات القدرات المعرفية، والاستعداد الدراسى ، والتحصيل.

(٣) مقاييس القدرات الفارقة التى أعدها إيليوت Elliot عام ١٩٩٠.

Differntial Ability Scales (DAS)

وتشتمل على اختبارات معرفية واختبارات تحصيل للأطفال الذين تتراوح



أعمارهم بين ٢,٥ ، ١٧ عامًا، وتُعد هذه المقـاييس تعديلاً لمقـاييس الذكاء البـريطانية British Ability Scales . وقد تم تنظيـم مقاييس القـدرات الفارقة فـى مجمـوعتين إحداهما البطارية المعرفية Cognitive Battery وتشتمل على مسـتويين أحدهما لأطفال ما قبل المدرسة والآخر لأطفال المدرسة الابتدائية، والمجـموعة الأخرى اختبارات تحصيل دراسى تشتمل على مهارات حسابية أساسية ، وهجاء، وقراءة كلمات.

وإلى جـانب هذه المقاييس الفـردية المهمـة توجد مـقاييس عـديدة أخرى لـذكاء الأطفال حديثى الولادة، مثل مقياس برازيلتون (عام ١٩٧٣).

Brazelton Neonatal Assessment Scale (BNAS)

وكذلك مقياس بيلى Bayley (١٩٦٩) لنمو الأطفال الذين تـتراوح أعمارهم بين شهريـن، ٣٠ شهراً Bayley Scale of Infant Development ، ويقيس الوظائف العقلية والحركية لهؤلاء الأطفال.

كما توجد مقايس فردية للأطفال من ذوى الإعاقـات وفئات خاصة، مثل مقياس كـولومـبيـا للنضج العـقلى (عـام ١٩٧٢) Columbia Mental Maturity Scale (CMMS) للأطفال الذيـن يعانون من إعاقـات جسمية، أو قـصور لغوى، أو فـقدان سمع، وتتراوح أعـمارهم بين ٣ ، ١٢ عامًا، ولا يتطـلب المقياس استجـابات لفظية أو مهارات حركية دقيقة. ومقياس إيلينوى للقدرات السيكولوجية اللغوية Illinois Test of Psycholinguistic Abilities (ITPA) وهو من المقاييس الحديثة نسبيًا التى تقيس العجز فى التعلم Learning Disabilities لدى الأطفـال الذين تتراوح أعـمارهم بين عـامين وعشرة أعوام . واسـتند المقـياس إلى المفاهيم المعاصرة فى تجهـيز المعلومات Human Information Processing حيث افترض أن إخفاق الطفل فى الاستجابة الصحيحة لمثير معين ربما يكون ناتجًا ليس فقط من نظام مـخرجات قـاصر (أى استـجابة) ، وإنما من مدخـلات قاصرة، ومن نظام تجهـيز معلومـات قاصر أو من أحدهمـا. وقد لاقى هذا المقياس اهتمامًا كبيراً من جانب المربين والمهتمين بهذه الفئات الخاصة وكذلك الباحثين.

ويلاحظ أن معظم هذه المقاييس تقتصر على الأطفـال الصغار ، وعددًا قليلاً منها يناسب الأطفال الأكبر سنًا والمراهقين . وبعض هذه المقاييس يهتم بالمهـارات غير اللفظية أو الذكاء غير اللفظى مثل المهارات الإدراكية الحركية. كما أنه يمكن أن نحصل باستخدام بعض هذه الاختبـارات على درجة كلية واحدة كـما فى مقياس بينيـه ، والبعض الآخر يمكننا من الحـصـول علـى عـدة درجات كـما فى مـقاييس ويكسلر . ويتطلب تـطبيق

بعض هذه الاختبارات فاحصًا مدربًا تدريبًا عاليًا مثل مقاييس ويكسلر وبينيه ، في حين يتطلب البعض الآخر قدر أقل من المهارة ، أو تطبق تحت إشراف اخصائي.

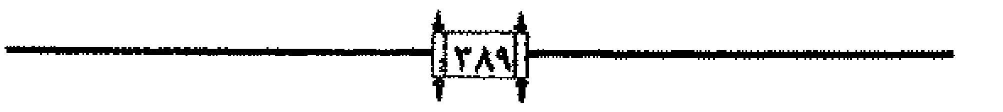
ثانيا : اختبارات الذكاء الجماعية :

سبق أن أوضحنا عند عرضنا لقياس الذكاء من منظور تاريخي نشأة حركة اختبارات الذكاء الجماعية بغرض انتقاء الجنود وتصنيفهم والتوسع بعد ذلك في استخدامها في مختلف مجالات الحياة المدنية. وتتميز هذه الاختبارات بإمكانية تطبيقها على أعداد كبيرة من الأفراد في وقت واحد مما يُقلل زمن العملية الاختبارية وما تتطلبه من نفقات مالية وإمكانات فنية على عكس ما تتطلبه اختبارات الذكاء الفردية، ولعل ذلك كان السبب الرئيسي لبناء اختبار الجيش ألفا Army Alpha Test، واختبار الجيش بيتا Army Beta Test ،واختبار التصنيف العام لأفراد القوات المسلحة (AGCT) ، والاختبار البديل له (AFQT) ، وهي الاختبارات الجماعية التي أشرنا إليها من قبل.

واختبارات الذكاء الجماعية تكون عادة اختبارات ورقة وقلم ويشتمل معظمها على مفردات اختيار من متعدد مما ييسر عملية التصحيح. وبذلك لا تتطلب هذه الاختبارات مواد مرتفعة الكلفة أو فاحصين على درجة عالية من المهارة لتطبيق الاختبار أو تصحيحه بموضوعية وسرعة، إذ يمكن تصحيحه باستخدام الحاسوب. غير أن استخدام هذا النوع من المفردات في كثير من هذه الاختبارات يجعل من غير الممكن اختبار الأطفال أو الأميين أو الذين لديهم إعاقات سمعية أو بصرية بها مما يتطلب اختبارات ذكاء جماعية غير لفظية . .

وبينما تُستخدم اختبارات الذكاء الفردية في تقييم وتشخيص المشكلات النفسية في المواقف الكلينيكية، فإن اختبارات الذكاء الجمعية تُستخدم في المدارس والجامعات والمصانع والمؤسسات العسكرية، كما تُستخدم بكثرة في البحوث التربوية والنفسية. ويمكن الإفادة من نتائج هذه الاختبارات في اتخاذ قرارات انتقاء العاملين Personnel Selection ،وأغراض التصفية Screening، والتوجيه المهني Vocational Guidance ، وذلك لتقييم القدرات العقلية، والمهنية، والاستعدادات الخاصة المتعلقة بمهن أو أعمال معينة.

غير أن هذه الاختبارات لا تسمح بتدخُّل القائمين بتطبيقها من حيث حفز المختبرين أو الثناء على استجاباتهم وهو ما يميز تطبيق اختبارات الذكاء الفردية. وهنا يُفترض أن المُختبرين لديهم دافعية مرتفعة وميل إلى التعاون دون حثهم على ذلك.



ونظراً لأن بعض المُختبرين ربما لا يكونون كذلك، أو تكـون لديهم مشكلات صحيـة أو خوف من الفشل أو قلـق ، فإنه يحتمـل أن يحصلوا على درجات أقـل مما يستـحقون. والحقـيقة أن الظروف الـطارئة التي ربما تـؤثر في أداء الفرد في الاخـتبار يصعب التعرف عليها في مواقف الاختبارات الجـماعية. لذلك ينبغي الحيطة في تفسير درجات هذه الاخـتبارات، واعتـبار هذه الدرجـات تقديرات تقـريبية وليـست درجات مطلقة.

ويُفـضل الحصـول على أكثـر من درجة واحـــدة للفـرد قـبل اتخـاذ قرارات أو استدلالات تتعلق بذكائه أو إمكاناته العقلية، وقد سبق أن ناقشنا ذلك في الفصل الثاني عندما تناولنا الشروط العلمية لتفسير درجات الاختبارات وتقديم نتائجها.

كما أن هـذه الاختبارات أقل فـائدة في التنبـؤ بالنجاح الأكاديـمي نظراً لأن نوع الذكاء المطلوب في الدراسة الأكاديمية متعـلق بنوعية الدراسة. فهذا النوع يؤكد بدرجة كبيرة الاسـتدلال اللفظي والعددي، بينما يؤكد الجوانب الاجتـماعية والعملـية للذكاء بدرجة أقل. ففـي المواقف الاكاديمية نهـتم بقدرة الفرد في الأداء الأكاديـمي، فالقدرة المرتفعـة على الاستدلال المجرد لدى الفرد ربـما تختلف اختلافًا كبـيراً عن القدرة أو الذكاء المطلوب للتعامل في المواقف الحياتية والتكيف مع هذه المواقف.

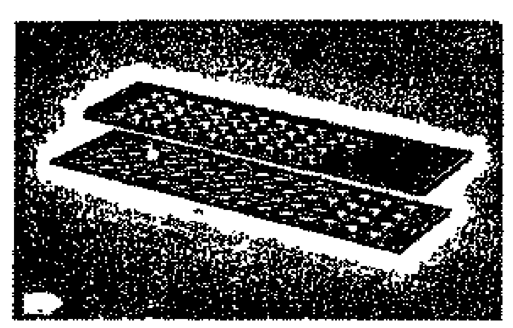
أنواع اختبارات الذكاء الجماعية:

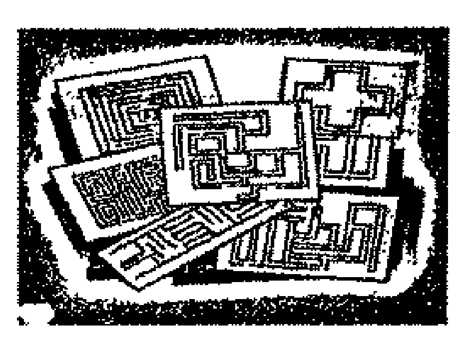
تتعـدد أنواع اختبـارات الذكاء الجـماعية وتـتباين تـباينًا ملحـوظًا في محـتواها وتكوينها وكيفية الإجابة عن مفرداتهـا، ولعل هذا يتضح بدرجة كبيرة عند إطلاعنا على المجلد الأخير للكتاب الـسنويّ للقياس النفسي Mental Measurement Yearbook الذي سبق أن أشرنا إليه (Buros, 1998) .

محتوى الاختـبارات: يمكن تصنيـف محتوى اختـبارات الذكاء الجمـاعية إلى نوعين : محتوى لفظي Verbal، ومحتوى غير لفظي Nonverbal ،فالمحتوى اللفظي يتطلب الـقدرة على القراءة أو على الأقـل فهم التعليـمات التي يُقدمـها القائم بعـملية التطبيق سواء قام بقراءتها أو توضيحها.

أما المحتوى غيـر اللفظي فلا يتطلب استخدام اللغة سواء فـي تطبيق الاختبار أو في الإجابة عنه، ومثال ذلك اختبار الجيش بيتا Army Beta الذي أشرنا إليه من قبل. وهذا النوع من الاختبارات يكون مناسبًا بل وضروريًا للأطفال الصغار والأميين وضعاف العقول والفئات الأقل حظًا الذين لم ينالوا قدراً مناسبًا من اللغة.

ويشتمل محتوى هذه الاختبارات غير اللفظية على مواد مصورة ، وأشكال ، ومتاهات، ومواد عيانية أخرى، كما يمكن أن تتطلب نوعًا من الأداء العملى Preformance كما فى مقاييس ويكسلر مثل بناء مكعبات، ومتاهات ، وألغاز مكانية. ويوضح شكل (٩ ـ ١١) التالى بعض هذه المواد غير اللفظية (Noll & Scannell, 1972)





شكل (٩ ـ١١) يوضح بعض المفردات غير اللفظية لاختبار المتاهات

أما مفردات الجانب اللفظى فى معظم اختبارات الذكاء الجماعية فتشتمل على: كلمات، وإكمال جمل، وحساب، ومتشابهات لفظية، ومتضادات لفظية، وسلسلات لفظية وعددية، وتمييز بين الفروق فى تسلسل من الكلمات، وأمثال ، وترتيب جمل، واستدلال لفظى، واسترجاع مؤجل، أى إجابة سلسلة من الأسئلة بعد قراءة قصة معينة. أى أن محتوى هذه الاختبارات يتعلق بالكلمات، والمعلومات العامة، والحساب، والاستدلال.

وعلى الرغم من أن الارتباط بين نوعى الاختبارات اللفظية وغير اللفظية ليس مرتفعًا بدرجة كبيرة، إلا أن مقدار الارتباط يسمح بتبرير استخدام المحتوى غير اللفظى فى الحالات التى تتطلب ذلك وبخاصة الأداء العملى. لذلك نجد أن كثيرًا من اختبارات الذكاء التى تُستخدم فى وقتنا الحاضر يشتمل على كل من المحتويين، وأحيانًا يكون محتوى الاختبار بطريقة تسمح بالحصول على نسبتى ذكاء إحداهما تتعلق بالجانب اللفظى والأخرى تتعلق بالجانب غير اللفظى، وتُمثلان معًا الذكاء العام مقاسًا بطريقتين مختلفتين.

تكوين الاختبارات :

على الرغم من أننا نحصل باستخدام بعض الاختبارات الجماعية على درجتين إحداهما للجانب اللفظى، والأخرى للأداء العملى إلا أن هاتين الدرجتين لا تنفصل إحداهما عن الأخرى، أى لا تدلان على قدرتين مختلفتين.

ولكن هناك اختبارات نحصل باستخدامها على مجموعة من الدرجات الفرعية التى تُمثل جوانب مختلفة للذكاء، وتسمى بطاريات اختبارات متعددة العوامل Multifactor Aptitude Batteries سوف نتناولها فى الفصل العاشر. وعادة تُصنف الاختبارات بحسب الصف الدراسى أو المستوى العمرى. غير أنه توجد اختبارات ذات مستويات مختلفة يناسب كل مستوى منها صفوفًا دراسية معينة، أى أنها تقيس الذكاء فى مستويات الاختبار المتعددة. وتُختار عينات التقنين بحيث يمكن المقارنة بينها من مستوى إلى آخر، وهذا يسمح بمقارنة القياسات التى نحصل عليها عبر الزمن.

فلكى نحصل على قياسات النمو العقلى بمفردات متسعة المدى فى صعوبتها تُبنى سلسلة من بطاريات متداخلة متعددة المستويات Multilevel Batteries. ويُختبر كل فرد بالمستوى المناسب له فقط، ويمكن إعادة اختباره فى الأعوام التالية بالمستويات الأخرى، كما يمكن استخدام هذه المستويات فى المقارنة بين مجموعات عمرية مختلفة، لذلك كان لابد من إحداث تداخل بين البطاريات المتسلسلة.

وتُصنف عادة مفردات اختبارات الذكاء الجمعية، ويتم تجميع مفردات كل صنف فى اختبار فرعى واحد، ومن ثم تُرتب مفردات الاختبار الفرعى ترتيبًا تصاعديًا بحسب مستوى صعوبتها. غير أن بعض الاختبارات تخلط أنواع المفردات بشكل حلزونى شامل Spiral Omnibus Forms حيث تُقدم مفردة من كل نوع (لفظى بأنواعه، أو غير لفظى مثلاً) ، وتكرر السلسلة ولكن بمفردات أكثر صعوبة. وهذا يسمح للفرد باستمرار الإجابة ولا تكون هناك ضرورة لتحديد زمن إجابة لكل اختبار فرعى.

كيفية الإجابة عن مفردات الاختبارات :

كثير من اختبارات الذكاء الجماعية تتطلب السرعة والدقة فى الإجابة Speed Tests وهذه تكون اختبارات موقوتة، وتتميز مفرداتها بالسهولة النسبية حيث يجيب الفرد عن عدد كبير منها فى وقت محدد. غير أن هناك بعض الاختبارات تعتمد على القوة Power Tests ، وهذه تكون مفرداتها متدرجة فى صعوبتها وليس لها زمن محدد. وفى الحقيقة لا يستطيع الإجابة عن جميع مفردات الاختبار الذى يتطلب السرعة إلا الأفراد النابهين، أما درجة الفرد فى اختبارات القوة فتعتمد على صعوبة المفردات التى يستطيع الإجابة عنها.

أمثلة لبعض اختبارات الذكاء الجماعية :

نظراً لتعدد أنواع اختبارات الذكاء الجماعية وتنوعها حيث يوجد عدد كبير من هذه الاختبارات تناسب جميع المستويات العمرية من الحضانة إلى سن الرشد، فإنه يصعب حصر هذه الاختبارات أو تعريف القارئ بها، وإنما يمكن الرجوع إلى الكتاب

السنوى للقياس النفسى (Buros , 1998) للاطلاع على أحدث هذه الاختبارات بالتفصيل، وكذلك أدلة الاختبارات فى البيئة العربية.

غير أننا سوف نقدم فى الجزء التالى أمثلة لبعض أنواع الاختبارات الأكثر استخدامًا فى قياس الذكاء، وعلى الرغم من أنها مقننة فى البيئة الأمريكية إلا أن بعضًا منها تم تعريبه وتقنينه فى البيئة المصرية وفى بيئات عربية أخرى.

اختبارات ذكاء جماعية لفظية متعددة المستويات:

(١) اختبار كاليفورنيا للنضج العقلى California Test of Mental Maturity (CTMM) عام ١٩٦٣:

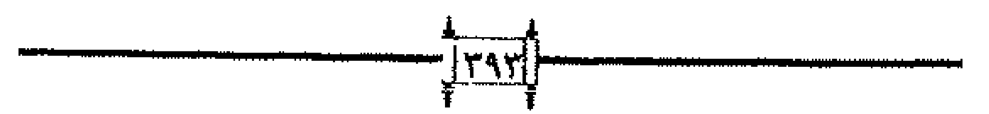
ويشتمل على سلسلة من الاختبارات الفرعية عددها ١٢ اختبارًا تقيس عمليات عقلية تتمثل فى الذاكرة، والعلاقات المكانية، والاستدلال المنطقى، والاستدلال العددى، ومفاهيم لفظية. والاختبار يتكون من عدة صيغ للقياس التتابعى فى مستويات ستة مخصص لكل منها كتيب اختبارى.

وهذه المستويات تبدأ من الحضانة إلى سن الرشد. ويمكن باستخدام الاختبار الحصول على ثلاث درجات كلية للجزء اللفظى والجزء غير اللفظى، والاختبار ككل. ويستغرق تطبيقه ما بين ٤٨ دقيقة ، ٨١ دقيقة للمستوى، لذلك اختُصر الاختبار فى صيغة أقصر.

(٢) اختبار القدرات المعرفية Cognitive Abilities Test (CAT) عام ١٩٧٨:

وهو اختبار متعدد المستويات حيث يشتمل على بطارية اختبارات للصفوف من الحضانة إلى الصف الثالث الابتدائى Primary Battery ،وبطارية متعددة المستويات Multilevel Battery للصفوف من الثالث إلى نهاية مرحلة التعليم الثانوى. وتتكون البطارية الأولى من صيغتين لكل مستوى، وكلاهما يشتمل على أربعة اختبارات فرعية هى: الكلمات الشفوية، والمفاهيم العلاقية، وعمليات عقلية متعددة، ومفاهيم كمية. ويستغرق تطبيق كل اختبار فرعى ربع الساعة تقريبًا فى أربع جلسات منفصلة، وتستخدم معايير نسبة الذكاء الانحرافية والمئينيات فى تفسير الدرجات.

أما البطارية الثانية فتشتمل على ثلاث بطاريات محتواها لفظى، وكمى، وغير لفظى على الترتيب. ويتضمن المحتوى اللفظى كلمات، وإكمال جمل، وتصنيف لفظى، ومتشابهات لفظية، ويتضمن المحتوى الكمى علاقات كمية، وسلاسل أعداد، وتكوين معادلات، وتعامل مع الأعداد وغيرها من الرموز الكمية، ويتضمن المحتوى غير اللفظى أنماط أشكال هندسية ومكانية، وهذه تتضمن تصنيف أشكال ، ومتشابهات



شكلية، وتكويـن أشكال، وكل من هذه المحتويات تقـاس باختبارات فرعيـة. وتعتمد معايير هذه البطاريـة ذات المستويات المتعددة على الدرجات المكافـئة للعمر، والرتب المئينية. والتساعيات المعيارية المكافئة للصفوف.

ومن الجدير بالذكر أن هذا الاختبار يعد تطـويراً للاختبار الشائع الاستخدام الذى أعده لورج، وثورنديك Lorge - Thorndike Test عام ١٩٦٣.

(٣) اختبارات أوتيس وليـنون للقـدرة العقلية العامة Otis - Lenon Mental Ability Tests عام ١٩٦٧:

وتتكون هذه الاختبارات من صيغة واحدة، وتشتمل على مفردات متنوعة بشكل حلزونى شامل Spiral - Omnibus لقياس القدرة العقلية العامة. وتوجد ستة مستويات للاختبارات تمتد من الصف الأول الابتدائى إلى نهاية مرحلة التعليم الثانوى، ويستغرق تطبيق الاختبارات مـن ٣٠ – ٤٥ دقيقة بحسب المستوى. ونحصل باستخدام الاختبار على درجة واحدة للفرد. وتعتمد المعايير على نسب الذكاء الانحرافية، والمئينيات، والتساعيات المـعيارية المقابلة للعمر والصف الدراسى، وكـذلك يمكن الحصول على الأعمار العقلية للمستويات الثلاثة الأولى.

(٤) اختـبارات الذكاء التشاركيـة بين المـدارس والكليـات Cooperative School and College Ability Tests , Series II (SCAT , Series II) عام ١٩٦٧:

وهذه الاختبارات من إعداد مركـز الخدمات الاختبارية ETS بالولايات المتحدة الأمريكية بالتعاون مع العديد من الخبراء التربويين، وهى سلسلة من الاختبارات متعددة المستويات تقيس الـقدرة العقلية العامة. وتُستخدم فى أربعة مـستويات تبدأ من الصف الرابع الابتـدائى. وتوجد صيغـتان متكافـئتان أو ثلاث لكل اخـتبار فى كل مـستوى. ويشتمل كل منها علـى ٥٠ مفردة تتعلق بالمتشابهات اللفظيـة، ٥٠ مفردة أخرى تتعلق بالمقارنات الكمية. ويمكن باستخدام هذه الاختبارات الحصول على ثلاث درجات كلية للجوانب اللفظية والكمية وللاختبار بعامة، وتُحوَّل هذه الدرجات إلى رتب مئينية.

(٥) اخـتـبار كـالمـان - وأنـدرسـون للـذكـاء Kuhlmann - Anderson Test (KAT) عام ١٩٦٣:

ويُعد من اختبارات الذكاء الجماعية،ويشتمل على ثمانية مستويات منفصلة لجميع الصفوف المدرسية، أى من الحضانة إلى نهاية مرحلة التعليم الثانوى، وتشتمل كل من هذه المـستويـات على اختـبارات متـعددة يحـتوى كل منـها على أنواع مـختلفـة من المفردات.

وهذه المجموعة من الاختبارات تُعد بمثابة بطارية متعددة المستويات Multilevel Battery نظراً لأنها تشمل مستويات متعددة من العمر أو الصف الدراسي. ومفردات الاختبارات في المستويات العمرية الدنيا محتواها لفظي ولا تتطلب إلا قدراً ضئيلاً من مهارات القراءة، وكلما ارتفع المستوى يزداد المحتوى اللفظي، ولكن يظل المحتوى غير اللفظي سائد. لذلك فإن هذه البطارية تُناسب الأطفال الصغار، وتُفيد أيضًا في قياس ذكاء من لديهم صعوبات في القراءة أو اتباع التعليمات اللفظية.

ويُعد هذا الاختبار من أقدم وأفضل اختبارات الذكاء الجماعية للأطفال حيث اعتنى ببنائه عناية فائقة ، وأجرى تعديله عدة مرات منذ عام ١٩٢٧ حتى طبعته الأخيرة السابعة.

والدليل على تميُّز هذا الاختبار أن كثيراً من مفرداته استُخدمت في اختبارات جماعية متنوعة أخرى. ويمكن الحصول على ثلاث درجات: لفظية، وكمية ، وكلية، وفي بعض المستويات يمكن التعبير عن هذه الدرجات بنسب ذكاء انحرافية ، وفي مستويات أخرى يعبِّر عنها بواسطة معايير المدى المئيني Percentile Bands لتقدير فترات الثقة في الرتب المئينية، كما سبق أن أوضحنا في الفصل السادس. ويتراوح زمن تطبيق الاختبار بين ٥٠، ٦٠ دقيقة.

(٦) اختبارات هنمون، ونلسون للقدرات العقلية

Henmon - Nelson Tests of Mental Ability (H - NT) عام ١٩٧٣ :

ويُعد هذا الاختبار أيضًا من الاختبارات الشائعة الاستخدام واعتُنى ببنائه بحيث يشمل أربعة مستويات بحسب الصفوف المدرسية أي الصفوف ٣ ـ ٦ ، ٦ ـ ٩ ، ٩ـ١٢ ، وهذه تُمثل الصيغة الأولى، والصفوف من الحضانة إلى نهاية الصف الثاني، وهذه تمثل البطارية الأولية. وقد أُعدت في طبعة حديثة اختبارات للمستوى الجامعي. وكل من المستويات الثلاثة للصيغة الأولى تشتمل على ٩٠ مفردة مرتبة ترتيباً حلزونيًا شاملاً Spiral Omnibus، وتتناول كلمات مختلطة ، ومتشابهات لفظية، وتصنيف لفظي، واستدلال لفظي، وسلاسل عددية، واستدلال حسابي، ومتشابهات شكلية، واتباع تعليمات. ويتراوح زمن تطبيق اختبارات هذه الصيغة ٣٠ دقيقة تقريبًا.

أما البطارية الأولية فتشتمل على ثلاثة اختبارات فرعية هي: اختبار استماع إلى معلومات عامة، واختبار كلمات مُصوَّرة، واختبار حجوم وأعداد. وهذه البطارية غير موقوتة، ولكن تطبيقها يستغرق عادة حوالي ٣٠ دقيقة.

وتعتمد معايير جميع الاختبارات على نسب الذكاء الانحرافية، والتساعيات المعيارية، والرتب المئينية المقابلة للعمر والصف الدراسي. وعلى الرغم من أن الاختبار يؤدي إلى درجة كلية واحدة يفترض أنها تمثل الذكاء العام، وهو ما تهدف الاختبارات لقياسه، إلا أنه يمكن الحصول على مجموعتين من المعايير يعتمدان على توزيع الدرجات الخام وفقًا لكل من العمر والصف الدراسي على الترتيب.

اختبارات ذكاء جماعية غير لفظية:

ذكرنا فيما سبق أن اختبارات الذكاء الجماعية غير اللفظية NonVerbal تستخدم في قياس ذكاء الأفراد الأميين، وبعض الفئات الخاصة التي لديها إعاقات ثقافية أو لغوية أو جسمية، والأفراد الذين تكون قدرتهم على القراءة محدودة، وكذلك الأطفال الصغار. ويمكن أن تكون هذه الاختبارات من النوع الذي يتطلب ورقة وقلم، أو يتطلب أداء مهام معينة أوضحناها من قبل. وهذه الاختبارات لا تعتمد على اللغة، بل يمكن تطبيق بعضها دون الحاجة إلى تعليمات لفظية.

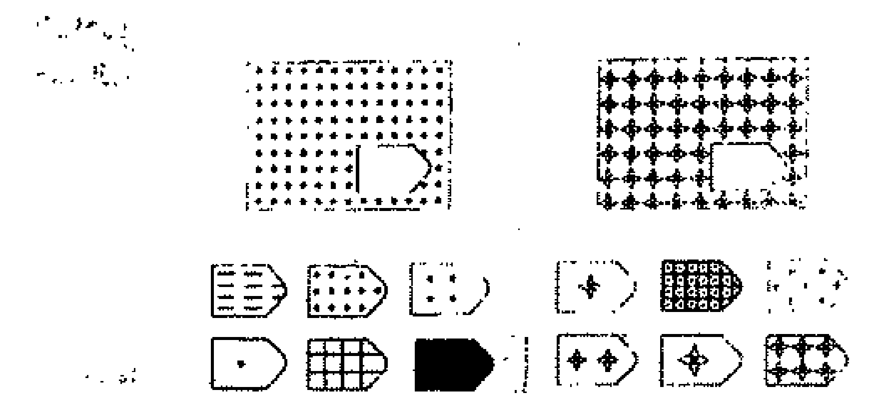
وفيما يلي أمثلة لبعض هذه الاختبارات الشائعة الاستخدام:

(١) اختبار المصفوفات المتتابعة لرافن

Raven Progressive Matrices عام ١٩٤٧:

وقد أعد عالم النفس رافن J. Raven وعالم الوراثة بنروس L. Penrose هذا الاختبار في إنجلترا. ويُعد من أكثر مقاييس الذكاء الجماعية غير اللفظية شيوعًا واستخدامًا في قياس القدرة العقلية العامة، وقد صُمِّم لتقييم ذكاء الأفراد المجندين دون أن تتأثر درجاتهم بالعوامل المتعلقة بالتعليم.

ويشتمل الاختبار على ٦٠ مصفوفة أو تصميم أحد أجزائه مقطوع، وعلى الفرد أن يختار الجزء المقطوع من بين بدائل معطاة عددها ستة، أو ثمانية. ومفردات الاختبار مصنفة في خمس مجموعات متسلسلة، كل منها تشتمل على ١٢ مصفوفة متزايدة الصعوبة. وتتطلب الإجابة إدراك المتشابهات، وإجراء تبديلات على الأنماط وغير ذلك من العلاقات المنطقية. ويوضح شكل (٩ ـ١٢) التالي مثالا لهذه المصفوفات (Jensen, 1980):



شكل (٩ ـ ١٢) يوضح مثالاً من مفردات اختبار المصفوفات المتتابعة

وللاختبار صيغ ثلاث إحداها صيغة معيارية، والثانية صيغة ملونة تناسب الأطفال في مرحلة ما قبل المدرسة، وضعاف العقول، والثالثة صيغة متقدمة أكثر صعوبة تناسب المستويات الأعلى من التعليم الابتدائي والمتوسط.

ويُعد الاختبار من نوع اختبارات القوة التي لا تتطلب زمنًا محددًا للإجابة، ولكنه يستغرق زمنًا يتراوح بين ١٥ ، ٣٠ دقيقة، ويمكن تطبيقه فرديًا أو جماعيًا بتعليمات شفوية بسيطة للغاية. ومعاييره المئينية تم تحديدها باستخدام عينات من تلاميذ المدارس البريطانية والراشدين.

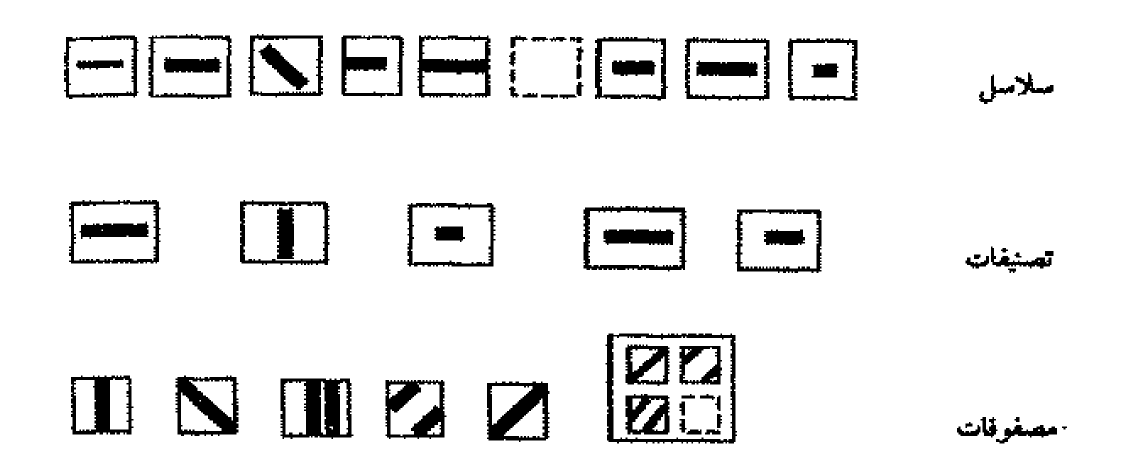
ويُعد هذا الاختبار من النوع الذي يطلق عليه الاختبارات غير المتحيزة للثقافة Culture - Fair Tests ، والهدف من هذه الاختبارات إتاحة فرص متكافئة للأفراد من ثقافات مختلفة في إجابتهم عن الاختبار. غير أنه تبين أنه ربما يصعب بناء مفردات اختبار يكون محتواها مستقلاً عن الخبرات المتباينة للأفراد من ثقافات مختلفة. كما أن هذه الاختبارات لا تتنبأ بدرجة جيدة بالتحصيل الأكاديمي أو الأداء في العمل، وصدقها في أغراض الانتقاء، والتسكين يكون منخفضًا بالمقارنة بالاختبارات الجماعية المعتادة.

(٢) اختبار الذكاء غير المتحيز للثقافة

IPAT Culture Fair Intelligence Test عام ١٩٥٧ :

وهذا الاختبار شائع الاستخدام في الولايات المتحدة الأمريكية، وقد أعد تحت إشراف كاتل R. Cattell في معهد قياس الشخصية والقدرات. ويشتمل الاختبار الذى يتكون من صيغتين متكافئتين على ثلاثة مستويات لكل صيغة كالتالي: المستويات العمرية ٤ ـ ٨ أعوام، والراشدين من ضعاف العقول و (٨ ـ ١٤) عامًا ، والراشدين من متوسطي الذكاء، وتلاميذ المرحلة الثانوية والراشدين المتفوقين. ويشتمل كل اختبار على أربعة اختبارات فرعية تقيس قدرة الفرد على إدراك العلاقات بين الأشياء، مثل

السلاسل ، والتصنيفات، والمصفوفات. ويوضح شكل (٩ ـ١٣) التالى مثالاً لبعض هذه المفردات :



شكل (٩ ـ ١٣) يوضح عينة من مفردات اختبار الذكاء غير المتحيز ثقافيًا

وبالإضافة إلى ذلك يوجد مقياس آخر ينقسم إلى أربعة اختبارات فرعية تقيس معلومات ثقافية وفهم لفظي. ويستغرق تطبيق الاختبار بعامة مدة زمنية تتراوح بين ٤٠ ، ٦٠ دقيقة. وقد وجد أن قيمة معامل ارتباط الاختبار بكل من مقياس ويكسلر ومقياس بينيه تتراوح بين ٠,٥٦ ، ٠,٨٥ (Tannenbaum , 1968) .

وقد أمكن التحقق من عدم تحيز الاختبار للثقافة بمقارنة معاييره التى تم إعدادها في كل من دول أوروبا الغربية والولايات المتحدة الأمريكية واستراليا، ووجد أن هناك تناظرًا فيما بينها.

(٣) اختبار الرسم: (Goodenough - Harris Drawing Test (G - HDT :

يُعد من الاختبارات غير اللفظية الجيدة، ويمكن تطبيقه فرديًا أو جماعيًا، ويطلب من الفرد رسم صورة رجل كامل، وتُقدَّر الدرجة على أساس تفاصيل الجسم والملابس، وتناسب الملامح والخصائص الأخرى، ولا يهتم بالنواحى الجمالية فى الصورة.

والاختبار غير موقوت، ولكن الفرد يستغرق عادة ١٥ دقيقة تقريبًا لرسم الصورة. وقد حدِّدت خصائص صورة الرجل التى تميز بين مختلف الأعمار، وتقدر لكل خاصية درجة واحدة، فمثلاً إذا رسم الفرد رأس الرجل دون تحديد تفاصيل ملامحه تُقدَّر له درجة واحدة، ومن ثم تُجمع الدرجات. وقد أُعدت معايير الاختبار



للأطفال الذين تتراوح أعمارهم بين ٣ ، ١٥ عامًا متضمنة الدرجات المعيارية ، والرتب المئينية لكل من الذكور والإناث.

وعلى الرغم من أنه كان يفترض أن الاختبار غير متحيز للثقافة، إلا أنه قد تبين أن عملية رسم رجل تتأثر تأثرًا ملحوظًا بالثقافة، ومع هذا فإن الاختبار يتميز بسهولة التطبيق وسرعته ولا يتطلب إلا ورقة وقلم، ونحصل باستخدامه على تقدير تقريبي سريع لذكاء الطفل. ويمكن الإفادة من هذا الاختبار بدمجه ضمن بطارية من اختبارات الذكاء لتأكيد المعلومات التى نحصل عليها نظرًا لأن تطبيقه بمفرده ربما يؤدى إلى نتائج غير دقيقة.

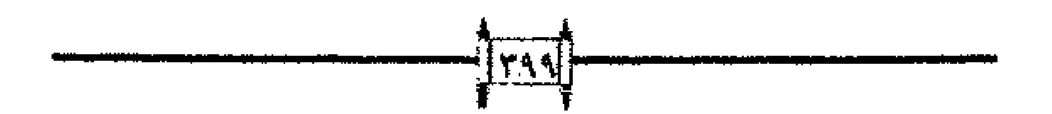
وقد أُجرى تعريب كثير من مقاييس الذكاء الجماعية والفردية وتقنينها في البيئة المصرية وفي بيئات عربية أخرى، ويمكن لمستخدمى الاختبارات والباحثين الرجوع إلى أدلة الاختبارات المنشورة في الوطن العربي لانتقاء الاختبارات والمقاييس المناسبة للغرض المطلوب.

تعقيب عام على مقاييس الذكاء:

يتضح مما سبق أن اختبارات الذكاء يمكن أن يُختبر بها كل فرد على حدة، أو تُطبق على مجموعة من الأفراد في وقت واحد وذلك لقياس القدرة العقلية العامة أي الذكاء العام. ولعل التساؤل الذى يترتب على ذلك هو كيفية اختيار الاختبار المناسب من بين هذين النوعين الرئيسيين. والحقيقة أنه لا توجد إجابة محددة عن هذا التساؤل، وإنما يمكن مراعاة بعض الاعتبارات في هذا الشأن.

فلكل من نوعي الاختبارات استخداماته ومزاياه وعيوبه، فاختبارات الذكاء الفردية تناسب بدرجة أفضل الأطفال الصغار لكى يُختبر كل منهم على حدة، وذلك لأن كثيرًا منهم لا يستطيع القراءة أو اتباع التعليمات التفصيلية، ومن السهل تشتيت انتباههم، مما يتطلب من الفاحص المتمرس توجيههم وحفزهم على الإجابة في إطار إجراءات مقننة للحصول على قياسات دقيقة.

لذلك يُوصى بعدم استخدام اختبارات الذكاء الجماعية في قياس ذكاء الأطفال الذين تقل أعمارهم عن خمسة أعوام، أى أطفال ما قبل المدرسة نظرًا لحاجتهم إلى الإشراف الفردي. أما الأطفال الذين يبلغون من العمر خمسة أو ستة أعوام فيمكن تطبيق اختبارات الذكاء الجماعية عليهم لقياس قدرتهم العقلية العامة، ولكن في مجموعات صغيرة قدر الإمكان، أي لا تزيد عن ١٥ طفل مثلاً.



ولكن نظراً لأن قيم معامل استقرار درجات اختبارات الذكاء الجماعية بالنسبة لأطفال الصفوف الأولى من المرحلة الابتدائية عبر الزمن تكون منخفضة، فإن استخدام هذه الدرجات في اتخاذ قرارات تربوية مهمة تتعلق بهؤلاء الأطفال ربما لا يكون مناسبًا.

لذلك إذا دعت الحاجة إلى قياس الذكاء العام لبعض الأطفال الصغار لأغراض معينة، فإنه يمكن استخدام اختبارات الذكاء الفردية، أو استخدام بعض الاختبارات الجماعية المناسبة لمستواهم النمائي والتي لا تتطلب إلا قدراً ضئيلاً من القراءة أو الكتابة، وإنما يضع الطفل علامة على ورقة الأسئلة مباشرة حيث إن الاختبار في هذه الحالة يتضمن مواد غير لفظية.

لذلك يفضل استخدام اختبارات الذكاء الجماعية وبخاصة بطاريات الاختبارات متعددة العوامل **Multi-Factor Batteries** في قياس ذكاء المراهقين والراشدين العاديين حيث إن التنبؤ باستخدام هذه الاختبارات يكون جيداً بنفس القدر إذا استُخدمت الاختبارات الفردية. كما أن هؤلاء لا يتأثرون كثيراً بمواقف الاختبارات الجماعية، ويعملون على توظيف أقصى إمكاناتهم في إجاباتهم.

كذلك تتميز هذه الاختبارات الجماعية بأنها قليلة الكلفة، ولا يستغرق تطبيقها وقتًا طويلاً، ولكنها لا تناسب المواقف الكلينيكية التي تتطلب عمليات التشخيص، حيث ينبغي استخدام اختبارات الذكاء الفردية لهذا الغرض.

لذلك فإن الاختبارات الفردية تتطلب جهداً كبيراً في بنائها بواسطة اختصاصيين في القياس النفسي، فإعداد المواد المتنوعة التي يشتمل عليها الاختبار، وإجراء عمليات تقنينه وإعداد معاييره، وتحديد وصياغة تعليماته، وطريقة تصحيحه، وتدريب الفاحصين يستغرق وقتًا طويلاً ونفقات مالية كبيرة.

وعمومًا، فإن اختيار أيٍّ من نوعيّ الاختبارات يعتمد على الهدف الأساسي من استخدامهما، فإذا كان الهدف تقدير قدرة الأفراد الذين يكون الموقف الاختباري الجماعي مناسبًا لهم من حيث التعليمات ومستوى الدافعية وما إلى ذلك فإنه يُفضل استخدام اختبارات الذكاء الجماعية. أما المواقف التي تتطلب معلومات تفصيلية أكثر دقة، وكذلك ملاحظات الفاحص، فإنه يُفضل اختيار أحد الاختبارات الفردية التي ناقشناها فيما سبق.

بعض القضايا المتعلقة بقياس الذكاء :

(١) بناء اختبارات الذكاء :

عرضنا في الجزء السابق وجهات نظر متعددة تتعلق بمفهوم الذكاء ونماذجه ونظرياته وأدوات قياسه، ولعل من المناسب أن نوضح الأساس المنطقي الذي تستند إليه اختبارات الذكاء بأنواعها المختلفة التي أشرنا إليها.

فاختبارات الذكاء بعامة تُعد اختبارات مقننة ويُتبع في بنائها خطوات مماثلة لخطوات بناء الاختبارات التحصيلية المقننة أو الاختبارات التحصيلية مرجعية الجماعة أو المعيار التي ناقشناها بإسهاب في الفصل الثامن. إذ ينبغي أولا تحديد النطاق السلوكي الذي يقيسه اختبار الذكاء ، وابتكار مفردات مناسبة لقياس مكونات هذا النطاق. وكتابة المفردات يُعد علما وفنا له أصوله وقواعده، وينبغي كتابة عدد من المفردات أكبر مما هو مطلوب، وتختلف تفاصيل عملية بناء هذه الاختبارات باختلاف الجماعة المرجعية التي يطبق عليها الاختبار، أي ما إذا كانت هذه الجماعة من المراهقين والراشدين أم من الأطفال.

فبعد تطبيق الاختبار على العينة المستهدفة من المراهقين والراشدين يجرى تحليل مفردات الاختبار كما سبق أن أوضحنا في الفصل السابع، حيث ينبغي تحديد النسبة المئوية من الأفراد الذين أجابوا إجابة صحيحة عن كل مفردة.

ولا يجب أن تكون معظم المفردات غاية في السهولة أو الصعوبة، فإذا أجاب عدد محدود من الأفراد عن المفردات إجابة صحيحة أو إجابة خطأ، فإن الاختبار لن يُميز بين الأفراد تمييزًا كافيًا. وإذا كان مستوى صعوبة كثير من المفردات متوسطًا، أي أن معامل الصعوبة ٠,٥٠ فإن الاختبار لن يُميز بين الحالات المتطرفة في توزيع الدرجات.

وكذلك ينبغي إيجاد معامل الارتباط بين درجات مفردات الاختبار والدرجات الكلية. فالمفردات التي أجيب عنها إجابة صحيحة يجب أن ترتبط ارتباطًا موجبًا بالدرجات الكلية، كما يجب أن تكون الارتباطات بين المفردات موجبة أيضًا. أما إذا كانت العينة المستهدفة من الأطفال، فإنه إلى جانب تحليل المفردات السابق يكون بناء الاختبار أكثر صعوبة، وذلك لأن ذكاء الأطفال يتوقع أن يزداد بزيادة العمر، حيث يجب أن تزداد نسبة الأطفال الذين يجيبون عن المفردات إجابة صحيحة بزيادة العمر.

لذلك سوف تكون هناك كثير من المفردات إما سهلة أو صعبة للغاية بالنسبة لأطفال مستوى عمرى معين، ولكى لا يحبط الطفل فإنه عادة يُوقَف تطبيق الاختبار إذا أخفق الطفل فى إجابة عدد محدد من المفردات المتدرجة فى صعوبتها فى مرات متتالية. ويُفترض عند تحديد هذه الخصائص السيكومترية للمفردات أن الذكاء عامل عام، غير أن هناك نماذج أخرى للذكاء كما رأينا فى مستهل هذا الفصل مثل نموذج ثيرستون أو جيلفورد يفترض عدداً من العوامل أو القدرات المستقلة. وفى هذه الحالة ينبغى مراعاة أن ترتبط المفردات المتعلقة بكل عامل أو قدرة فيما بينها ارتباطاً مرتفعاً نسبياً، وترتبط بغيرها من المفردات المتعلقة بالعوامل الأخرى ارتباطاً منخفضاً.

ومعظم اختبارات الذكاء تُحدد النسبة المئوية من المفردات التى يجيب عنها الأفراد إجابة صحيحة، وهذا يُعد أحد محكات تقييم الذكاء. غير أن ثورنديك (Thorndike, 1968) قد أشار إلى أن قياس الذكاء لا يقتصر على قياس العُلو أو الارتفاع Altitude. فإذا أجاب فرد عن عدد من المفردات المتكافئة في صعوبتها أكبر مما أجاب عنه فرد آخر فى اختبار غير موقوت، فإن الفرد الأول يمكن اعتباره أكثر ذكاء من الفرد الثانى، وعندئذ يقال أن لديه وسع أو اتساع Breadth. وكذلك إذا استطاع فرد أن يجيب عن مفردات أكثر سهولة من فرد آخر في اختبار موقوت، فإن الفرد الأول يمكن اعتباره أكثر ذكاء من الفرد الثاني نظراً لأنه أكثر سرعة فى إجابته.

والحقيقة أن سرعة الإجابة تُعد من المتغيرات الشائكة، وذلك لأن الأفراد يختلفون في مقدار الزمن الذى يحتاجونه لحل مشكلات معينة، فالبعض ربما يترك المفردات الصعبة بالنسبة له دون إجابة بمجرد مواجهتها، وبعض آخر ربما يترك المشكلات التي كان من الممكن حلها بمزيد من المثابرة، ومع هذا، فإن السرعة كمؤشر للذكاء أصبحت تهتم بها الدراسات السيكولوجية فى الآونة الأخيرة، فالأفراد الذين يمكنهم تجهيز المعلومات ومعالجتها بسرعة أكبر من غيرهم ربما يستطيعون بمرور الزمن تكوين ذخيرة من المعلومات الضرورية، وبالتالى يحصلون على درجات مرتفعة فى اختبارات الذكاء.

ومن المشكلات المهمة الأخرى فى عملية بناء الاختبارات هو أن تكون السمة المراد قياسها أحادية البُعد Unidimensional. فالبُعد الأحادى يُمكّننا من تفسير معنى الدرجة الناتجة. فإذا كان الأداء فى الاختبار يعتمد على بُعدين، فإنه من الممكن أن يحصل فردان على الدرجة نفسها بطرق مختلفة، إذ ربما يكون الفرد متميزاً فى أحد الأبعاد وضعيفاً فى البُعد الآخر، والعكس بالنسبة للفرد الثانى.

لذلك ينبغى التحقق من فرضية أحادية البُعد باستخدام أساليب إحصائية مناسبة مثل التحليل العاملى الذى سبق أن أوضحناه، وذلك عند بناء اختبارات الذكاء أو غيرها من الاختبارات المقننة.

(٢) بعض محددات الذكاء والعوامل المؤثرة فيه :

اتضح لنا مما سبق أن معظم اختبارات الذكاء روعى فى بنائها أن يكون متوسط نسب الذكاء أو الدرجات المحوّلة ١٠٠ ،فنجد أن ثلثى أفراد الجماعة المرجعية تقريبا تحصل علـى درجات تتراوح بين ٨٥ ، ١١٥ (ذكاء متوسط أو قريب من المتوسط)، ١١٪ تقريبًا تكون درجاتهم أعلى من ١٢٠ (متفوقين أو موهوبين) ، ١,٦ ٪ يحصلون على درجات أعلى من ١٤٠ (متفوقين جدًا)،بينما يحصل ٨,١ ٪ تقريبًا على درجات تنحصر بين ٣٠ ، ٧٩ (فئات الضعف العقلى)، وذلك وفقًا للمنحنى الاعتدالى الذى أوضحناه فى الفصلين الثالث والسادس.

غير أن توزيع الدرجات يختلف إلى حد ما من اختبار إلى آخر. ويعتمد هذا الاختلاف علـى ثبات درجات الاختبار، والسمات أو العوامل التى يقيسها، وطبيعة الجماعة المرجعية التى يُطبق عليها الاختبار وتُستمد منها معاييره.

ويمكن أن يرجع تباين الأفراد والجماعات فى الذكاء إلى مصادر متعددة، مثل العوامل الوراثيـة التى تبدو أهميتها فى هـذا المجال من البحوث المتعددة التى تناولت هذه العوامل بالدراسة وبخاصة البحوث التى قارنت بين ذكاء الآباء والأبناء، وبين التوائم المتطابقة والتوائم المتشابهة، وكذلك العوامل البيئية التى تلعب دورًا مهما فى هذه الاختلافات فى الذكاء وبخاصة فى المراحل المبكرة من نمو الطفل.

تأثير العوامل الوراثية فى الذكاء :

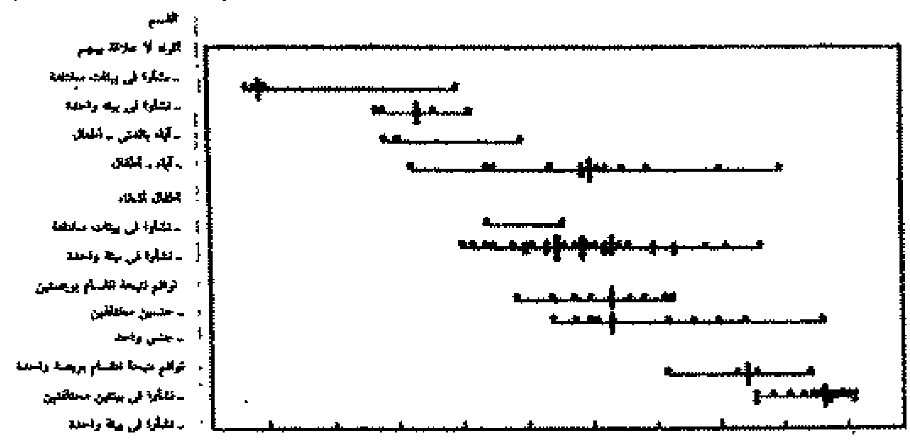
لقد بيَّن جنسن (Jensen, 1969) بجامعة كالفورنيا فـى مقالته الشهيرة «إلى أى حد يمكننا رفع الذكاء والتحصيل الدراسى How Much Can We Boost IQ and Scholastic Achievement» ، أن السبب الرئيسى فى عدم فاعلية البرامج التربوية التعويضية التى تقدم للأطفال الأقل حظًا فى الولايات المتحدة الأمريكية أنها ركزت على الخبرات البيئية كمحدد للنمو العقلى.

وقدَّم بيانات تؤيد وجهة نظره مؤداها أن ٨٠٪ تقريبًا من الفروق فى نسب الذكاء تُعزى إلى الفروق الوراثية، فى حين أن العوامل البيئية تسهم بنسبة ضئيلة فى هذه الفروق، وذكر أن المستوى الأول للقدرة Level I Ability المتعلق بعمليات الترابط Associative يكون توزيعه متعادلاً بين مختلف الفئات الاجتماعية، بينما المستوى الثانى للقدرة Level II Ability المتعلق بالتفكير التصورى Conceptual يتوزع توزيعًا متباينًا بين الفئات الاجتماعية والاقتصادية المنخفضة والمتوسطة، حيث يتميز الأطفال من المستويات الاجتماعية،والاقتصادية المتوسطة عن الأطفال من المستويات المنخفضة فى هذه القدرة الثانية.

أى أنه يرى أن الاختلاف بين هؤلاء الأطفال يرجع إلى اختلافهم فى نمط قدرات التعلم، مما يتطلب تقديم أساليب تعليمية مختلفة لتنمية القدرات المتوافرة لدى كل من المجموعتين. وقد أكد أيزنك (Eysenck, 1971) فى كتابه «الجدل حول نسبة الذكاء The IQ Argument» وجهة نظر جنسن Jensen.

غير أن جنسن فى كتابه «التحيز فى القياس العقلى Bias in Mental Testing» الذى صدر عام ١٩٨٠ عاد وذكر أن أسباب هذه الفروق متعددة، وتحديد هذه الأسباب يعد من المهام المعقدة التى تتطلب بحوثًا متعددة التخصصات، وبخاصة فى العلوم البيولوجية والسلوكية والقياس النفسى.

وفى دراسة أكثر اتساعًا قام إلينماير وكيملنج (Elenmeyer - Kimling, 1963) بتلخيص نتائج ٥٢ دراسة أجريت فى ثمانية دول، على مدى النصف الأول من القرن الحالى حول مشكلة الوراثة فى الذكاء الإنسانى يوضحها شكل (٩ ـ ١٤) التالى:



شكل (٩ ـ ١٤) يوضح نتائج ٥٢ دراسة تتعلق بالوراثة فى مجال الذكاء الإنسانى

ويتضح من شكل (٩ ـ ١٤) أن هناك اتساقًا فى النتائج على الرغم من اختلاف الدراسات فى تكوين العينات من حيث المستوى الاجتماعى والاقتصادى والجنسية وأنواع الاختبارات المستخدمة. إذ يبدو أنه كلما اقتربت العلاقات الوراثية بين الأفراد كانوا أكثر تماثلا فى الذكاء.

ولعل قيم الارتباطات الموضحة تُعد الحدود الدنيا لهذه القيم، نظرًا لأن الدراسات اعتمدت على بيانات مستمدة من الأطفال وليس من الراشدين، ولم تفصل بين الذكور والإناث مما أدى إلى تقديرات لتأثير العوامل الوارثية أقل من حقيقتها،

فوسيط القيم المبنية في شكل (٩ ـ ١٤) ١, ٤١ ويقترب كثيراً من قيمته المتوقعة. إذا اعتبرنا الوراثة هي المحدد الوحيد لدرجات اختبارات الذكاء. غير أن تأثير العوامل البيئية يتضح من قيم الارتباط بالنسبة للأفراد الأقارب الذين نشأوا معًا التى تبدو أعلى من القيم بالنسبة للأفراد الأقارب الذين نشأوا فى بيئتين مختلفتين على الرغم من ثبوت العوامل الوراثية.

كما يمكن ملاحظة الفروق بين الأفراد غير الأقارب الذين نشأوا معًا أو فى بيئتين مختلفتين. وبذلك ربما يوضح شكل (٩ ـ ١٤) النزعة العامة لدور كل من الوراثة والبيئة كمحددات لدرجات اختبارات الذكاء.

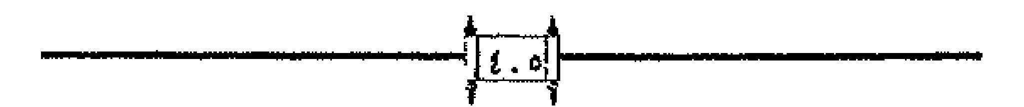
وفى إطار الجدل حول رفع نسبة الذكاء ينبغى أن نُميز بين ذكاء الفرد، والدرجة التى يحصل عليها فى اختبار للذكاء. ففى معظم الدراسات التى أجريت حول إمكانية رفع نسبة الذكاء اتُخذت الدرجة المرتفعة فى الاختبار مؤشراً على ارتفاع القدرة العقلية العامة، على الرغم من أن هذا موضع تساؤل.

فالتحاق الأطفال بدور الحضانة واكتسابهم بعض الخبرات يساعدهم فى الإجابة عن مفردات اختبارات الذكاء بدرجة أفضل قليلاً، ولكن إذا استُخدمت مفردات مختلفة ربما لا يستطيعون الإجابة عن هذه المفردات مما قد يدل على عدم وجود تأثير متسق لهذه الخبرات فى القدرة العقلية العامة.

فبعض علماء النفس عرّف الذكاء بأنه ما تقيسه اختبارات الذكاء. والارتفاع الطارئ فى نسبة الذكاء يمكن تفسيره إلى حد ما فى ضوء هذا التعريف. فالطفل الذى يلتحق بدور الحضانة يكتسب اتجاهات ومهارات ربما تساعده فى أن يصبح أكثر تمرسًا فى التعامل مع الاختبار، وذلك لزيادة ثقته بالكبار المختلفين عن والديه، مما يجعله أكثر تقبلاً للفاحص عندما يكرِّر تطبيق الاختبار عليه من أجل المتابعة.

كما أن التفات الطفل المتكرر إلى المواد التى يشتمل عليها الاختبار، مثل المكعبات والصور واللوحات ربما يزيد من ألفته بها، مما يمكّنه من التعامل معها بدرجة أفضل.

وهذان العاملان ربما يُسهمان فى رفع نسبة ذكائه عدة نقاط. ولكن هل يعنى ذلك أن الطفل أصبح أكثر ذكاء؟ فلقد ذكرنا فيما سبق أن الذكاء لا يقاس قياسًا مباشرًا. وقياسه بطريقة غير مباشرة يعتمد على تقييم استجابات الطفل لسلسلة من المواقف المقننة، أى مفردات الاختبار. والهدف من اختبار الذكاء هو تقدير درجة تدل على قدرة الطفل على التعامل مع هذه المواقف بحيث يمكنه توظيف هذه القدرة فى التعامل مع جميع المواقف المماثلة.



ولكي يتحقق هذا الهدف يجب أن يشتمل الاختبار على عينة غير متحيزة وكافية من المفردات. فإذا تدرب الطفل على كيفية إجابة مثل هذه المفردات، هل أداؤه عندئذ يتعلق بعينة غير متحيزة من المفردات، أم أن هذه العينة تكون مصطنعة وتدل على أن الطفل ربما يؤدي أداء جيدًا فقط إذا واجه مواقف مصطنعة مماثلة.

فمن المعلوم أن معظم الأفراد يُوظفون قدرًا ضئيلا من قدراتهم العقلية. لذلك ربما يكون من المفيد تقديم خبرات إثرائية للطفل في حل المشكلات والفحص والبحث وما إلى ذلك. وعندئذ يمكن الإفادة من اقتراح جنسن Jensen بأن الذكاء تحدده العوامل الوراثية، وذلك بالعمل على تنويع الخبرات الإثرائية وجعلها ملائمة لنمط تعلم الطفل لكي يوظّف إمكاناته العقلية الموروثة إلى أقصى حد ممكن.

تأثير العوامل البيئية في الذكاء :

إن التساؤل حول أثر البيئة الاجتماعية في الإسراع بالنمو العقلي للأطفال كان ولا يزال مثار جدل وحوار وأبحاث ودراسات مستمرة. فقد تبين من كثير من البحوث أن هناك علاقة وثيقة بين ذكاء الأطفال والمستوى الاجتماعي والاقتصادي لآبائهم (McNemar , 1942 ; Harrel, 1945; Harrell, 1945 ; Eckland, 1965; Tyler, 1965 ; Ginitis 1973)

ولكن يجب عدم المبالغة في تأكيد هذه العلاقة، فالأطفال الذين تقع نسبة ذكائهم في المدى الممتد من التأخر العقلي إلى العبقرية إنما هم لآباء من كل نمط مهني. كما تبين من نتائج بعض الدراسات أن الأسر الذكية بحكم الوراثة (أي بيولوجيًا) قد تستغل ذكائها في توفير بيئات أكثر إثراء، وبذلك تعمل على رفع مستوى ذكاء أفرادها.

فمثلاً تبيّن من كثير من الدراسات أن الأطفال في المؤسسات الخاصة بضعاف العقول تميل نسبة ذكائهم إلى الانخفاض بزيادة أعمارهم، أما إذا وضعوا في ملاجئ للأيتام لكي يتيسر لهم في هذه الملاجئ الاستثارة العقلية على الرغم من عدم تهيئة مناخ يسوده الأمن والثقة لا يُحْدث أي انخفاض في نسبة ذكائهم، ولكن إذا انتقلوا إلى بيوت تبنّ طيبة ترتفع أحيانًا نسبة ذكائهم، أو على الأقل يتوقف الانخفاض فيها.

ويرى البعض أنه يمكن رفع نسبة الذكاء عندما يُحوَّل الأطفال المتخلفون عقليًا من الفصول العادية إلى الفصول الخاصة، إلا أن نتائج البحوث التي أجريت في هذا

المجال لم تؤكد هذه الزيادة، في حين أن البرنامج التعليمي المُعد إعدادًا جيدًا ربما يُيسّر تحصيل هؤلاء الأطفال ويزيد من توافقهم الشخصي والاجتماعي. وعموما لا توجد إجابة محددة عن مدى تأثير العوامل البيئية في الإسراع في نمو الذكاء.

غير أن تحليل هذه العوامل المؤثرة في النمو العقلي يبيّن أن الذكاء هو نتاج العوامل الوراثية والبيئية، إذ بينت كثير من الدراسات أنه لا يمكن تنمية الذكاء أبعد من حدود الإمكانات الفطرية البيولوجية. ومن بين هذه دراسة تيرمان Terman ما بين عامي (١٩٤٩ - ١٩٥٩) المتعلقة ببحث الانتقال الوراثي للذكاء، فقد وجد أن الأطفال يميلون إلى أن يشبهوا آباءهم في الذكاء. ولكنه فسر هذه النتيجة بأن الآباء من ذوي الذكاء المرتفع ربما يوفرون بيئة أكثر إثارة لأبنائهم مما يؤدي إلى تفوقهم، أي أن الإسراع في النمو العقلي يكون نتيجة لتفاعل العوامل الوراثية والبيئية. وبالطبع من الصعب تحديد القدر الذي يسهم به كل من هذه العوامل في النمو العقلي.

(٣) ثبوت نسبة الذكاء:

هل تظل نسبة ذكاء الفرد ثابتة في مختلف سني حياته ؟

والحقيقة أنه لا توجد إجابة قاطعة لهذا السؤال، نظرًا لصعوبة تحديد نسبة كاء الفرد الواحد بشكل دقيق خلال مراحل حياته، ومع ذلك يمكن القول بأن نسبة ذكاء الفرد (متوسط الذكاء) تظل ثابتة نسبيًّا في مختلف مراحل نموه بوجه عام مع تغير ربما يتراوح في مدى ١٠ درجات.

ففي دراسة طولية أجراها هونزيك Honzik، وماركفارلين Markfarlane، وألين Allen (١٩٤٩) اعتنوا بتصميمها لبحث استقرار نسب الذكاء، حيث قاموا بتتبع ٢٥٢ طفلاً من الميلاد إلى الرشد. وقد اختُبر الأطفال باختبار ستانفورد - بينيه في مختلف مراحلهم العمرية، وقد تبين أنه كلما تقدم الطفل في العمر استقرت نسبة ذكائه. فمثلاً وُجد أن قيمة معامل الارتباط بين نسب الذكاء في المدى العمري ٤ - ٦ أعوام ٠,٦٢ بينما قيمته في المدى العمري ٨ - ١٤ عامًا ٠,٨٥ . أما درجات الاختبار في مرحلة ما قبل المدرسة كانت قليلة الفائدة في التنبؤ بالدرجات في المراحل التالية، بل كانت عديمة الفائدة في مدة زمنية فاصلة قدرها عام واحد.

وفي خلال سني الدراسة في المدى العمري ٦ - ١٨ عامًا تذبذبت نسبة ذكاء حوالي ٦٠٪ من العينة ١٥ نقطة أو أكثر، ومن بين كل ثلاثة أطفال تذبذبت نسبة ذكاء

أحدهم ٢٠ نقطة أو أكثر، ومن بين كل عشرة أطفال تذبذبت نسبة ذكاء أحدهم ٣٠ نقطة أو أكثر، ١٥٪ من أفراد العينة تذبذبت نسبة ذكائهم أقل من ١٠ نقاط.

وعمومًا ، فإن تذبذب نسبة الذكاء من تطبيق إلى آخر للاختبار تقل كلما تقدم الطفل في العمر، وهذه النتيجة لها أهمية تربوية حيث تُبين أن درجات اختبارات الذكاء العام في الصفوف المبكرة من المحتمل ألا تناظر الدرجات في الصفوف التالية، مما يتطلب تكرار تطبيق هذه الاختبارات في الصفوف المبكرة.

لذلك لا ينبغي استخدام نسبة الذكاء في وصف الطفل أو تصنيفه في نمط معين، لما لذلك من آثار نفسية وتعليمية واجتماعية على مستقبل حياته. فنسبة الذكاء تمثل أداء الطفل في وقت معين، وربما تبدو إمكانات الطفل أعلى أو أقل في الأعوام التالية.

غير أن تغير نسبة الذكاء يبدو أنه يرتبط بمستوى الأسرة من حيث تعليم الوالدين، والحالة الاجتماعية والاقتصادية. فالطفل الذى نسبة ذكاؤه مرتفعة ولكن مستوى تعليم والديه منخفضًا وفي مهن لا تتطلب مهارة، ربما تنخفض نسبة ذكائه كلما تقدم في العمر إلى أن يصل إلى مرحلة الرشد. وعلى العكس فإن الطفل الذى نسبة ذكاؤه منخفضة ولكن ثقافة والديه ومستوى تعليمهما متميزًا ، فإنه ربما تزداد نسبة ذكائه في مراحل نموه التالية (Duncan , Featherman and Duncan , 1972).

غير أنه ينبغي أيضًا مراعاة تأثير العوامل المتعلقة بالشخصية والدافعية في الذكاء، وبخاصة أنه ربما يختلف هذا التأثير باختلاف نوع الطفل.

ولا يزال هناك أيضًا جدل وعدم اتفاق حول استقرار ذكاء الراشدين. فقد اختبر ويكسلر (Wechsler , 1955) عينة مستعرضة من الراشدين ووجد أن الجوانب اللفظية للذكاء تزيد حتى سن ٣٠ عامًا ثم تبدأ في الانخفاض التدريجي، غير أن أسلوب المعاينات الذى استخدمه يشوبه بعض القصور نظرًا لاختلاف مستوى تعليم المجموعات العمرية المختلفة. فمستوى تعليم الأفراد الأصغر سنًا كان أعلى من مستوى تعليم الأفراد الأكبر سنًا، وربما أدى ذلك إلى هذه الفروق.

كما قام بيلي(Bayley , 1955) بدراسة طولية استنتج منها أن نمو الذكاء يستمر حتى يبلغ الفرد ٥٠ عامًا. غير أن الاستنتاج الذى يمكن الاستناد إليه هو أن الوظائف العقلية العامة لا تنخفض انخفاضًا آليًا بتقدم العمر. فكما ذكرنا يمكن أن تعمل العوامل البيئية للراشدين أيضًا على رفع أو خفض أدائهم العقلي، ولكن إذا أخذنا بعين الاعتبار

الظروف الصحية لراشد عمره ٥٠ عامًا ، فإن قدرته العقلية ربما تكون مساوية لقدرته عندما كان عمره ٢٥ عامًا (Jarvik et al . , 1973) .

ويبقى الآن السؤال التالي :

ما السن التي يتوقف فيها النمو العقلي ؟ أي ما السن التي يكتمل فيها النضج العقلي للفرد؟

فلقد حاول كثير من علماء النفس دراسة هذه المشكلة بالأساليب الإمبيريقية، حيث انتقوا اختبارات ذكاء معينة، وقاموا بتطبيقها على أفراد من أعمار مختلفة ، وقارنوا بين درجات الاختبارات لكل مجموعة متحدة في العمر الزمني والمجموعات الأخرى، ووجدوا استمرارًا في منحنى النمو حتى أعمار معينة . والاتجاه العام في الدراسات المختلفة يذهب إلى أنه قلما يحدث تغيير في النمو العقلي بعد سن الثامنة عشرة .

ولكن تظل الإجابة عن السؤال محاطة باعتبارات كثيرة في البحث عن أثر العوامل الاجتماعية والاقتصادية للعينة التي تطبق عليها اختبارات الذكاء، ومدى تمثيل العينة لجميع أفراد المجتمع، ومدى تنوع هذه الاختبارات بحيث تُظهر الاختلافات في القدرات أو الاستعدادات الخاصة . ولعل التمييز الذى اقترحه كاتل Cattell بين الذكاء المائع Fluid Intelligence ، والذكاء المتبلور Crystalized Intelligence، والذى أشرنا إليه في مستهل هذا الفصل يلقى ضوءًا على هذه المشكلة .

خلاصــة

من هذا يتضح أن مجـال الذكاء الإنسانى يُعد من المجالات الـشائكة التى نالت قدرًا كـبيـرًا من اهتمـام علماء الـنفس وخبـراء القيـاس بل والعلمـاء من فروع الـعلم المختلفـة، ولا يزال الحوار والجدل قائمًا حـول تعريف الذكاء، وأساليب ومنـهجيات قياسه، وكيفية الإفادة من نتائج مقـاييس الذكاء، والتغلب على المشكلات الناجمة عن سوء استخدام هذه النتائج فى المجالات التربوية والاجتماعية.

لذلك اتجهت بحوث الذكاء فى الآونة الأخيـرة وجهة جديدة، حيث اعتبرت أن تقييم الذكاء بالأسـاليـب التى تناولناها فى هذا الفصل يُعد تقييـمًا إستاتيكيًا؛ وذلك لأن هذا التقـييم لا يهتم كـثيرًا بإمكـانية إفادة الفرد مـن التعليم وبخـاصة بالنسبـة لحالات الضعف العقلى والأفراد الهامشيين، كما أن هذا التقييم يستند إلى مفهومًا ضيقًا للذكاء. ومن هنا حـاولت بعض الدراسـات الإفادة من التطورات الـتى حدثت فى علـم النفس المعرفى فى تطوير أساليب تـحليل العمليات المعرفية وتحديد مكوناتهـا وقياسها، كما حاولت تطوير أساليب جديدة لقياس إمكانات التعلم Learning Potential لدى الأفراد بحيث يكون التقييم ديناميكيًا وليس إستاتيكيًا كما هو الحال فى القياس التقليدى للذكاء الإنسانى.

وقـد أوضح هيـود Haywood، وبـراون Brown، وينجنـفلد Wingenfeld (١٩٩٥) أن التقييم الدينامـيكي للذكاء يؤكد تقييم العمليـات وليس النواتج ، ومحاولة تحديد العقبـات التى تحول دون التعلم الفاعل ، وكذلك محـاولة تحديد الشروط التى تسهم فى الأداء الجيد، والتمييز بين الأداء والإمكانات الفردية، كما أن هذا التقييم يؤكد تكامل العملية الاختبارية مع عملية التعليم.

لذلك ينبغي مراعاة أن الذكاء يُعد مفهومًا أكثر اتساعًا مما تشير إليه نسبة الذكاء، فهو يتضمـن القدرة العامة على التكيف فى مواقف الـحياة الاجتماعية الفعلـية، ويتميز الأفراد بالتنوع وتعدُد القدرات، مما يتطلب تعدُد جوانب القياس.

ولقد تناولنا فيما سبق بعض النماذج والنظريات التى اعتبرت الذكاء قدرات عقلية متـعددة ، غير أن هـذه القدرات ترتبط ارتـباطًا وثيـقًا بالدافعـية والاتجاهـات والأبعاد الأخرى ذات العلاقة.

لذلك لا يجب الارتكان إلى درجة كلية واحدة للذكاء فى اتخاذ قرارات مهمة تتعلق بالفرد، حيث إن هذه الدرجة ينبغى النظر إليها على أنها مدى من الدرجات قد يكون متسعًا أو ضيقًا وفقًا لثبات درجات الاختبار.

وتُعبِّر هذه الدرجة عن ذكاء الفرد فى وقت تطبيق الاختبار، ولكنهما ربما تتغير بمرور الزمن، وهذا التغير يمكن التحقق منه بتطبيق اختبارات أخرى، أو إعادة تطبيق الاختبار بفواصل زمنية معينة. وعند تفسير نسبة الذكاء ينبغى ربطها بخلفية الفرد من حيث الفرص المتاحة له للتعلم، ومهاراته اللغوية، وحفز الأسرة له على التعلم. كما أنه نظرًا لعدم توافر أدلة كافية حول التأثير النسبى لكل من الوراثة والبيئة فى الذكاء حيث إن كلاهما يؤثر بلا شك بدرجة معينة، فإنه ينبغى العمل على إثراء البيئة التعليمية للأطفال والمراهقين فى المنزل والمدرسة ومواقف العمل، وتصميم أساليب تعليمية وتدريبية تمكِّن الأفراد مهما اختلفت مستويات قدراتهم وتباينت سمات شخصياتهم من الإسهام الفاعل فى إثراء حياتهم وتنمية مجتمعهم.

وهذا يتطلب التحوُّل من القياس مرجعى الجماعة أو المعيار الذى يهتم بمقارنة أداء الفرد بأقرانه، إلى القياس مرجعى المحك الذى يُفيد فى تشخيص مهارات واحتياجات الأفراد التربوية والمهنية والاجتماعية بهدف تحسين وتطوير هذه المهارات لديهم، وقد سبق أن ناقشنا هذا النوع من القياس بإسهاب فى الفصل الثامن.

فالأفراد لا يختلفون فقط فيما بينهم فى القدرة العقلية العامة، وإنما أيضًا فى استعداداتهم الخاصة المتعلقة بمجالات تعليمية أو مهنية معينة، مما يدعو إلى الاهتمام بقياس وتقييم هذه الاستعدادات الخاصة، والعمل على الإفادة منها وتنميتها عن طريق تزويد الأفراد بالخبرات الثرية التى تُسهم فى إبرازها بالقدر الذى تسمح به إمكانات كل منهم، وهو ما سوف نتناوله بالتفصيل فى الفصل التالى.

الفصل العاشر

قياس الاستعدادات الخاصة

* مقدمة
* قياس الاستعدادات الخاصة من منظور تاريخى
* مفهوم الاستعداد وعلاقته بمفهوم القدرة
* تأثير العوامل الوراثية والبيئية فى الاستعدادات الخاصة
* استخدامات اختبارات الاستعدادات الخاصة
* أنواع اختبارات الاستعدادات الخاصة
* بعض الافتراضات الأساسية فى قياس الاستعدادات الخاصة
* أساليب بناء اختبارات الاستعدادات الخاصة
* إجراءات تحليل العمل
* اختبارات الاستعدادات الخاصة
* بطاريات الاستعدادات المتعددة
* اختبارات استعدادات تتعلق بمجالات نوعية محددة
* الاستعداد الميكانيكى
* الاستعدادات النفسية الحركية
* الاستعداد الكتابى
* الاستعداد الموسيقى
* الاستعدادات الفنية
* الابتكارية
* الابتكارية ومفهوم التفكير التباعدى
* قياس قدرات التفكير الابتكارى
* اختبارات التفكير الابتكارى
* مشكلات قياس التفكير الابتكارى
* خلاصة

مقدمة :

عرضنا فى الفصل السابق المفاهيم والنماذج النظرية المتعلقة بالذكاء الإنسانى، ومنهجيات وأساليب قياسه، وبعض محدداته الوراثية والبيئية وما يرتبط بهما من مؤثرات متعددة، وقد أوضحنا أن الذكاء هو القدرة العقلية العامة حيث اعتبرها بعض العلماء قدرة عامة واحدة تنطوى تحتها كافة الأنشطة العقلية، فى حين اعتبرها آخرون مجموعة من المكونات العاملية أو القدرات المتعددة التى يحدد تفاعلها مستوى القدرة العقلية العامة لدى الفرد، أى أن هذه القدرات اعتبرت بمثابة «عوامل الذكاء».

غير أن هناك تجمعات معينة لقدرات خاصة تبدو أهميتها فى أداء مهام محددة ويمكن الإفادة منها فى التنبؤ بالسلوك. وعادة يطلق على هذه التجمعات من القدرات الخاصة « الاستعدادات Aptitudes » .

فالاستعداد يشير إلى استطاعة الفرد اكتساب سلوك معين أو مهارات معينة إذا أتيحت له ظروف ملائمة أو تدرب مناسب. والذكاء العام أو القدرة العامة وما تنطوى عليه من قدرات متعددة ليست كافية لتعرف إمكانات الفرد وطاقته العقلية. فاختبارات الذكاء العام التى قدمناها فى الفصل السابق تبيِّن فاعليتها وجدواها بدرجة أكبر فى التنبؤ بالنجاح الأكاديمى بعامة، ولكن اختبارات الاستعدادات الخاصة Special Aptitudes تستخدم فى التنبؤ بالنجاح فى مجال دراسى أو تدريبى أو مهنى أكثر تحديداً، مثل النجاح فى دراسة اللغات الأجنبية أو الرياضيات العالية، أو التمكن من أداء أعمال ومهارات كتابية أو ميكانيكية أو فنية أو غير ذلك.

وقد سبق أن ميزنا فى الفصل الثامن بين اختبارات الاستعدادات الخاصة واختبارات التحصيل الدراسى، واختبارات الذكاء العام، وأوضحنا أنها تقع على متصل واحد، ويتباين موقعها عليه وفقًا لنوعية وشمول الخبرات التى تتطلبها فى إطار مفهوم القدرات النامية أو المستثمرة.

فلكى يحدد الفرد إمكاناته وقابلياته بصورة أكثر شمولاً ينبغى أن يتعرف على استعداداته الخاصة إلى جانب قدرته العقلية العامة، مما يتطلب توافر اختبارات مناسبة تقيس هذه الاستعدادات، إذ ربما تكون القدرة العقلية العامة لفردين مختلفين متقاربة، ومع هذا فإن كليهما يختلفان اختلافًا ملحوظًا فى نوعية استعداداتهما الخاصة المتعلقة بمجالات دراسية متنوعة أو أعمال ومهام وظيفية ومهنية معينة.

لذلك سوف نتناول في هذا الفصل مفهوم الاستعداد والعوامل المؤثرة فيه، وأنواع اختبارات الاستعدادات الخاصة واستخداماتها وأساليب بنائها، وأمثلة متعددة لبطاريات الاستعدادات، واختبارات تتعلق بمجالات نوعية محددة مثل الاستعداد الميكانيكي، والاستعدادات النفسية الحركية، والاستعداد الكتابي، والاستعدادات الفنية، ونوضح بمزيد من التفصيل مفهوم الابتكارية واختبارات التفكير الابتكاري، وما يتعلق بقياس الابتكارية من مشكلات.

قياس الاستعدادات الخاصة من منظور تاريخي :

إن تطور قياس الاستعدادات الخاصة من منظور تاريخي يرتبط ارتباطًا وثيقًا بتطور حركة القياس العقلي التي أوضحناها في الفصل السابق.

فقد ذكرنا أنه منذ قيام الحرب العالمية الأولى حتى عام ١٩٣٠ شهد مجال القياس العقلي نمواً متسارعًا حيث تزايد عدد الاختبارات الجماعية المقننة، واتسع نطاق استخداماتها، وصدر العديد من الكتب والمقالات والبحوث المتعلقة بموضوعات القياس. غير أنه بدأت تبرز تساؤلات متعددة سواء من المشتغلين ببناء الاختبارات أو من مستخدميها تتعلق بمحدودية هذه الاختبارات، وبخاصة اختبارات الذكاء وأوجه قصورها في قياس قدرات وإمكانات الأفراد.

وقد أدى ذلك إلى توسيع مجالات الاختبارات العقلية، وإجراء العديد من الدراسات والبحوث حول طبيعة القدرات العقلية والإفادة من نتائجها في تطوير هذه الاختبارات وغيرها من أدوات التقويم. ولعل قيام الحرب العالمية الثانية قد ساعد في تشجيع بناء اختبارات تستخدم في أغراض التوجيه التربوي والمهني، وبخاصة اختبارات الاستعدادات.

كما ساعد في ذلك البرامج الحكومية في الولايات المتحدة الأمريكية المتعلقة بالمجالات المهنية المختلفة مثل الزراعة والتجارة والصناعة والتشييد، وكذلك التوسع في برامج التعليم الثانوي وتنويعه مما تطلب تطوير برامج منظمة وفاعلة للإرشاد التربوي لطلاب هذا التعليم. وكذلك تطوير أساليب دقيقة لتصنيف وتسكين ملايين الأفراد المجندين.

وقد أسهم ذلك كله في دفع حركة بناء أنواع متعددة من اختبارات الاستعدادات للوفاء بهذه الاحتياجات الملحة، مثل اختبارات الاستعدادات الميكانيكية والكتابية والموسيقية والفنية.

وقد أدى زيادة التوجه نحو استخدام أساليب التحليل العاملى إلى تطوير طرق بناء اختبارات الاستعدادات الخاصة، حيث ساعدت هذه الأساليب فى التوصل إلى عدد من العوامل أو القدرات النوعية المستقلة نسبيًا حيث أمكن قياسها ببطاريات استعدادات متعددة منذ عام ١٩٤٥، وبخاصة فى المجالات العسكرية لاستخدامها فى انتقاء قائدى الطائرات والمعدات العسكرية والفنيين، وأعدت بعد ذلك بطاريات أخرى للأغراض التربوية والمهنية والإرشاد وانتقاء الأفراد وتصنيفهم.

واستمر استخدام أساليب التحليل العاملى فى دراسة التكوين العاملى للاختبارات المتوافرة وتطوير اختبارات جديدة استخدمت فى البرامج الاختبارية واسعة النطاق التى انتشرت فى الولايات المتحدة الأمريكية لأغراض متعددة، وكذلك اختبارات الاستعداد الدراسى، واختبارات القبول بالجامعات، واختبارات انتقاء الطلاب للالتحاق بالكليات الفنية والعملية والمهنية، وكذلك اختبارات انتقاء الطلاب للمنح الدراسية.

ومما ساعد بدرجة كبيرة فى تطوير واستخدام كثير من هذه الاختبارات وغيرها تأسيس مركز الخدمات الاختبارية بولاية نيوجرسى الأمريكية Educational Testing Services (ETS) عام ١٩٤٧، وكذلك البرنامج الاختبارى للكليات الأمريكية American College Testing Program الذى تأسس عام ١٩٥٩. وقد أدى التطور الكبير الذى حدث فى نظريات القياس ونماذجه منذ الستينيات إلى توجه جديد فى بناء أنواع متعددة من اختبارات الاستعدادات وإجراء عمليات تكافؤ درجاتها استنادًا إلى النظرية المعاصرة فى القياس التى أصبحت تعرف بنظرية الاستجابة للمفردة الاختبارية Item Response Theory (IRT) كما ذكرنا فى مواضع أخرى.

وهناك دراسات جادة يقوم بها علماء النفس بالقوات المسلحة الأمريكية فى وقتنا الحاضر لبحث إمكانية استخدام التصنيف المتتابع للقدرات الخاصة بالاستعانة بالحاسوب Computer Sequential Ability Testing فى مختلف أغراض الانتقاء والتدريب فى ضوء أسس ومبادئ هذه النظرية، مما يساعد فى تشخيص وتوجيه هذه القدرات الخاصة أثناء عمليات التدريب.

ويعتمد هذا التصنيف على برنامج اختبارى يُدار بواسطة الحاسوب تُحددً فيه مجموعة متسلسلة من المفردات الاختبارية لكل فرد، حيث يُختبر بمفردة تلو الأخرى، ويتخذ قرار تصنيف مبدئى عقب إجابته عن كل مفردة. فإذا أُجرى التصنيف فى حدود قدر معين من الخطأ المحدد مسبقًا لا يستمر الفرد فى الاختبار، أما إذا زاد مقدار خطأ التصنيف عما هو محدد يستمر الفرد فى إجاباته عن المفردات التالية إلى أن يتم التصنيف فى ضوء المحك المحدد للخطأ، أو إلى أن يُختبر الفرد بجميع المفردات المتوافرة.

ولعل هذا يماثل إجراءات تطبيق اختبار ستانفورد ـ بينيه ، فيما عدا أن إجراءات التطبيق والتصنيف تُدار آليًا بواسطة الحاسوب، ويتم تقدير استعداد الفرد باستخدام نظرية الاستجابة للمفردة الاختبارية المشار إليها.

كما أدى ذلك إلى التوجه نحو بناء نظرية علمية للقدرات المعرفية تُحدِّد وتُصنِّف جميع القدرات المعرفية الأساسية والفرعية بطريقة أكثر اتساقًا، وتُقدم أساليب دقيقة لقياس هذه القدرات، وتوضح كيفية إسهامها في الأنماط المختلفة للسلوك المعرفي والتعلم والأداء.

مفهوم الاستعداد وعلاقته بمفهوم القدرة:

لعل من المناسب في مستهل هذا الفصل أن نلقي مزيدًا من الضوء على مفهوم الاستعداد Aptitude وعلاقته بمفهوم القدرة Ability لكي يتسنى استخدام كل من المفهومين استخدامًا مناسبًا.

فقد اتضح لنا من الفصل التاسع أن الذكاء أو القدرة العقلية العامة تنطوي على تجمُّع معين من العوامل أو المكونات. غير أن معظم مقاييس الذكاء تركز على نسبة الذكاء بعامة (IQ) وليس على تقييم الجوانب النسبية للقوة والضعف في هذه المكونات كل على حدة. ولعل هذا هو ما يميز بين نسبة الذكاء وتقييم الاستعدادات الخاصة.

وسوف يتضح عند مناقشتنا لاختبارات الاستعدادات الخاصة أنها تتميز بعمق محتواها وارتباطه ارتباطًا وثيقًا بالاستعداد المراد قياسه، وهذا بالطبع لا يتحقق في اختبارات الذكاء. كما أن هناك استعدادات خاصة، مثل الاستعدادات الفنية، والاستعدادات الكتابية، والمهارات الحركية أو العضلية ، وغيرها يندر أن تتناولها معظم أساليب تقييم القدرة العقلية العامة.

والاستعدادات الخاصة تتضمن تجمُّعًا من القدرات الخاصة التي تعد ضرورية لأداء مهمة معينة أو تسلسل من المهام، وتجمُّع هذه القدرات بالنسبة للاستعدادات يماثل تجمُّع العوامل أو المكونات بالنسبة للذكاء العام. فالاستعداد يُفسَّر بواسطة ما تعلم الفرد أداءه من خلال توسُّط قدرته الخاصة.

ويمكن توضيح ذلك بالقدرة على التمييز بين طبقات الصوت Pitches، فبعض الأفراد لديهم قدرة مرتفعة على التمييز بدقة بين طبقات الصوت المختلفة، ويمكن قياس مدى دقة واتساق هذا التمييز، وهنا يبدو دور التعلم في تمكين الفرد من تسمية أو تصنيف طبقة معينة من طبقات الصوت الذي يسمعه نظرًا لأن هذه الإمكانية ليست موروثة.

وإذا بحثنا عن استعداد يتضمن هذه القدرة الخاصة المتعلقة بالتمييز بين طبقات الصوت نجد أنه الاستعداد الموسيقي. غير أن هذا الاستعداد يتضمن تجمعًا من القدرات الخاصة المتعلقة به، مثل إدراك أنماط الإيقاع، وذاكرة النغم وغيرها إلى جانب القدرة الخاصة السابقة. كما أن هناك مستوى آخر للاستعداد يسمى «الاستعداد من الدرجة الثانية Second Order Aptitude » وهو استعداد متعلق بالعمل JobAptitude. وهذا الاستعداد أكثر شمولاً من تجمع هذه القدرات الخاصة التى تمثل استعدادًا منفردًا ومتجانسًا نسبيًا مثل الاستعداد الموسيقي. فالاستعداد المتعلق بالعمل يمكن أن يشتمل على مجموعة متكاملة من الاستعدادات المترابطة التى تسهم فى الأداء الموسيقى.

فإذا تحدثنا عن قدرة الفرد ليكون موسيقارًا، فإننا ننظر إلى جميع المتطلبات أو المهام التى ربما يقوم بها الفرد أو يعرفها أو يفضلها وما إلى ذلك، ونضمها معًا لكى تكوّن الاستعداد المتعلق بالعمل. فمستوى الاستعداد الموسيقى المرتفع لدى الفرد يعد مظهرًا واحدًا من مظاهر الاستعداد الأوسع المتعلق بالعمل كموسيقار. وتحديد المكونات المتعددة التى يشتمل عليها هذا الاستعداد يتطلب إجراء تحليل للعمل Job Analysis إلى جانب ملاحظة الأداء الفعلى فى العمل.

من هذا يتضح أن الاستعداد الخاص لا يُعد عاملاً تكامليًا أو وحيدًا، وإنما يعد تنظيمًا معقدًا لمكونات متعددة. فالفرد ليس لديه قدرة كامنة واحدة، وإنما لديه قدرات كامنة متنوعة، ويتطلب أداء أى نشاط أو عمل معين كثيرًا من هذه القدرات الكامنة.

لذلك فإن بناء اختبار الاستعداد يعتمد على تحليل دقيق لمكونات الأداء المراد قياسه أو التنبؤ به. وبعض القدرات الكامنة يمكن أن تتداخل فيما بينها أو تكون مترابطة، فبعض المظاهر التى تُحدد القدرة الكامنة لفرد للنجاح فى مهنة معينة ربما تُحدد أيضًا إمكانية نجاحه فى مهنة أخرى (Wesman , 1973).

تأثير العوامل الوراثية والبيئية فى الإستعدادات الخاصة :

لا يستطيع أحد أن ينكر أن العوامل الوراثية تلعب دورًا مهما فى التعريف بحدود إمكانات الفرد. فقد اتضح لنا فى الفصل السابق أنه فى بداية القرن العشرين كان هناك تأييد من جانب كثير من علماء النفس لفكرة أن الذكاء وغيره من الاستعدادات موروث، وأن العوامل البيئية تلعب دورًا محدودًا فى نموها. غير أن تطور نظريات التكوين العقلى والبحوث التى أجريت بعد ذلك جعلت العلماء يولون العوامل البيئية اهتمامًا أكبر فى تنمية وإثراء الاستعدادات. وقد أكدت الدراسات المعاصرة تفاعل كل من العوامل الوراثية والعوامل البيئية فى الإسهام فى نمو الاستعدادات.

لذلك فإن اختبارات الاستعدادات الخاصة تقيس الأداء بجانبيه المتفاعلين: الموروث والمكتسب. وبهذا تكون اختبارات الذكاء أو القدرة العقلية العامة مرتبطة باختبارات الاستعدادات أكثر من ارتباطها باختبارات التحصيل، إذ يمكن اعتبار اختبارات الذكاء نوعًا خاصًا من اختبارات الاستعدادات وتقيس القدرة العقلية العامة أو الطاقة العقلية. وعلى الرغم من أن هناك عناصر مشتركة بينهما، بل وتسمى اختبارات الذكاء عادة اختبارات الاستعداد الأكاديمي Scholastic Aptitude وبخاصة إذا استُخدمت هذه الاختبارات في التنبؤ بالتحصيل الأكاديمي، إلا أن لكل منهما استخداماته الخاصة.

استخدامات اختبارات الاستعدادات الخاصة:

إن كثيرًا من الاختبارات السيكلوجية التي تستخدم في أغراض التنبؤ لا تفترض حصول الفرد على تدريب مناسب، كما أن بعضًا منها لا يقيس الجوانب المعرفية. أما اختبار الاستعداد فينبغي أن يقيس مهارة معرفية، ويستخدم أساسًا في التنبؤ أو الانتقاء أو التسكين.

فاختبارات الذكاء العام ربما تكون مناسبة لتقييم القدرة العقلية العامة للطالب التي تُمكنه من النجاح في كثير من البرامج الأكاديمية، أما اختبارات الاستعدادات الخاصة فإنها تستخدم في التنبؤ بالنجاح في تخصص أكاديمي أو مهني معين، مثل الهندسة أو الفنون أو الأداءات العملية المتنوعة أو المحاماة.

ويُعد انتقاء الأفراد Selection من أهم استخدامات اختبارات الاستعدادات. ويعتمد ذلك على تحديد درجة قطع Cutoff Score في الاختبار أو درجات قطع متعددة Multi Cutoff Scores إذا استخدمنا بطارية من الاختبارات. وينبغي أن ترتبط درجات الاختبار أو الاختبارات بمحك أداء مستقبلي متعلق بالاستعداد الذي يقيسه كل منها ارتباطًا مرتفعًا، وأن ترتبط فيما بينها ارتباطًا منخفضًا. وتتأثر درجات القطع بالسياسات التربوية أو المهنية، والاعتبارات الأخلاقية والاجتماعية. فتحديد درجة قطع مرتفعة للقبول بالكليات مثلاً سوف يقلل من الإنفاق الحكومي على التعليم العالي، ولكن هذا التوفير ينعكس أثره سلبًا على بعض فئات المجتمع، وبخاصة المستويات الاجتماعية والاقتصادية المنخفضة.

وقد سبق أن ناقشنا في الفصل الخامس الأخطاء الناجمة عن قرارات الانتقاء، وكيفية استخدام تحليل الانحدار في التنبؤ بدرجات المحك باستخدام الاختبارات التنبؤية، أي اختبارات الاستعدادات. وأوضحنا في سياق عرضنا للصدق المرتبط بمحك وبخاصة الصدق التنبؤي بعض المفاهيم، مثل نسبة النجاح ونسبة الانتقاء وجداول التوقع، وكيفية توظيفها في أغراض الانتقاء.

وتُستخدم اختبارات الاستعدادات الأكاديمية Scholastic Aptitudes Tests بكثرة في أغراض الإرشاد والتوجيه التربوى. فالمرشد التربوى فى المدرسة يهتم بمعرفة ما إذا كان التلميذ لديه صعوبات فى القراءة أو الرياضيات أو اللغات الأجنبية، وهذا التشخيص يساعد المعلم فى اقتراح خطة العلاج.

فمثلاً فى المرحلة الابتدائية يمكن استخدام اختبارات التأهب Readiness Tests للتعرف على الأطفال الذين لديهم صعوبات فى تعلم مواد دراسية معينة، وبخاصة القراءة والحساب. وفى المرحلة الثانوية يمكن استخدام اختبارات الاستعداد لتعلم اللغات الأجنبية فى التعرف على الطلاب الذين لديهم المهارات الضرورية لتعلم هذه اللغات،وتحليل أخطاء الإجابات فى هذه الاختبارات يمكن أن يساعد فى تحديد أفضل الأساليب العلاجية.

كما أنه يمكن باستخدام هذه الاختبارات التعرف على الطلاب الذين يتميزون باستعدادات مرتفعة فى المجالات العلمية والفنية والموسيقية مثلاً ، بحيث يمكن تصميم برامج إثرائية لهؤلاء الطلاب فى هذه المجالات.

أما اختبارات الاستعدادات المهنية Vocational Aptitudes Tests فتستخدم فى التنبؤ بالنجاح فى مهن أو أعمال معينة أو فى برامج تدريب على هذه الأعمال دون أن يتركز الاهتمام على الذكاء أو الاستعدادات الأكاديمية. وباستخدام هذه الاختبارات يمكن للمرشد أو المسئول عن التسكين أن يقدم معلومات للأفراد عن المهن أو برامج التدريب المناسبة لهم، أو أن يتخذ قرارات تسكينهم فى هذه المهن أو البرامج.

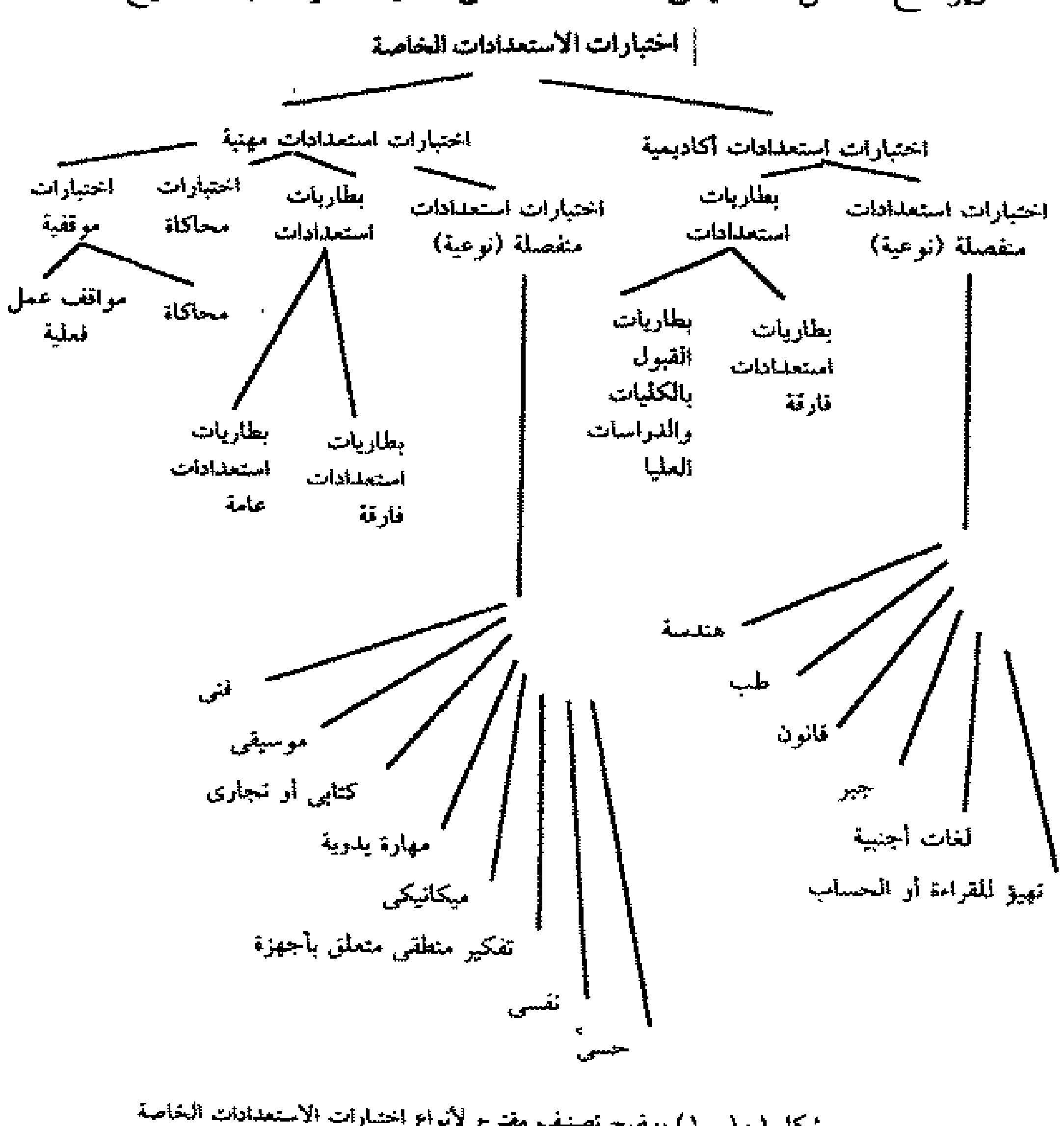
وكذلك تفيد هذه الاختبارات فى إرشاد الطلاب، وبخاصة فى مرحلة التعليم الثانوى عن نوع البرامج التعليمية المناسبة لاستعداداتهم والتى يمكنهم الالتحاق بها بعد انتهائهم من هذه المرحلة، مثل البرامج الأكاديمية، والبرامج التطبيقية التجارية أو الصناعية أو الزراعية وغير ذلك، وتوجيههم التوجيه المهنى المناسب.

ومن هذا يتضح أن بعض اختبارات الاستعدادات الخاصة يُستخدم كأدوات وصفية، وبعضها يستخدم كأدوات تنبؤية. فاستخدام هذه الاختبارات فى الإرشاد يهدف لوصف القدرات الخاصة للفرد أو الأفراد، وتعريف كل منهم بجوانب قوته وضعفه لكى يتخذ فى ضوئها قرارات تتعلق بمستقبله الدراسى أو المهنى. أما استخدام اختبارات الاستعدادات الخاصة فى التنبؤ فالهدف منه تقديم معلومات تفيد المسئولين عن انتقاء الأفراد سواء للالتحاق بالكليات أو بالمهن والأعمال المختلفة فى اتخاذ قرارات تتعلق بهؤلاء الأفراد، مثل اختبارات القبول بالكليات أو المعاهد المهنية والفنية، واختبارات انتقاء المتقدمين للوظائف الشاغرة فى مجالات العمل المختلفة.

أنواع اختبارات الاستعدادات الخاصة :

توجد أنواع متعددة من اختبارات الاستعدادات الخاصة في المجالات الأكاديمية والمهنية بعضها يقيس قدرة خاصة واحدة، وبعضها يقيس قدرات خاصة متعددة تشكل في مجملها استعدادًا خاصًا أو أكثر، وتُعد بمثابة بطارية من الاختبارات. وعلى الرغم من أن كثيرًا من هذه الاختبارات تتطلب إجابتها ورقة وقلما، إلا أن بعضًا منها يقيس أداء عمليًا أو حركيًا معينًا، وبعضًا آخر منها يتضمن عينات من المهام المتعلقة بعمل معين.

ويوضح الشكل التخطيطي (١٠ ـ ١) التالي تصنيفًا مقترحًا لهذه الأنواع:



شكل (١٠ ـ ١) يوضح تصنيف مقترح لأنواع اختبارات الاستعدادات الخاصة

(من تصميم المؤلف)

ويتضح من شكل (١٠ ـ ١) أنه يمكن تصنيف اختبارات الاستعدادات الخاصة في أربعة أنواع رئيسة نوجزها فيما يلي :

اختبارات استعدادات نوعية :

ويقيس كل منها قدرة خاصة أو استعداداً خاصاً واحداً في المجالات المهنية، مثل الاستعداد الحسي، أو النفسي الحركي، أو الميكانيكي، أو الفني، أو مثل التهيؤ للقراءة، أو استعداد اللغات الأجنبية، أو الهندسة، أو القانون، وغير ذلك من المجالات الأكاديمية. ويمكن أن تكون هذه الاختبارات مستقلة بذاتها أو ضمن بطارية من اختبارات الاستعدادات. وتفيد هذه الاختبارات في أغراض الانتقاء والتسكين، ولكنها محدودة الفائدة في أغراض الإرشاد. ويمكن أن تكون هذه الاختبارات من النوع الذي يتطلب ورقة وقلم، أو يتطلب أداء عملياً باستخدام أجهزة معينة.

لذلك فإن هذه الاختبارات ربما تطبق فردياً أو على جماعات محدودة العدد مما يجعل عملية التطبيق تستغرق كثيراً من الوقت.

بطاريات اختبارات الاستعدادات :

وتشتمل هذه البطاريات على عدد من الاختبارات المنفصلة التي تقيس مجموعة من القدرات الخاصة التي ينطوي عليها استعداد معين أو أكثر. ويمكن الحصول على درجة لكل من هذه القدرات أو المكونات، أو يمكن ضم جميع الدرجات للحصول على درجة كلية واحدة.

كما يمكن رسم صفحة نفسية (بروفيل) Profile للفرد باستخدام الدرجات التي حصل عليها في كل من الاختبارات التي تشتمل عليها البطارية. ولكن ينبغي مراعاة أن تكون معايير جميع الاختبارات مستمدة من جماعة مرجعية واحدة، وبذلك تُعطي قدراً كبيراً من المعلومات في أقل زمن ممكن. وبطاريات اختبارات الاستعدادات يمكن أن تكون فارقة Differential، أو عامة General، وتستخدم البطاريات الفارقة في مرحلة التعليم الثانوي لكي يتعرف الطالب على جوانب القوة والضعف في استعداداته، ويسترشد بذلك في تخطيط مستقبله التعليمي والمهني. غير أنه في مرحلة التعليم الابتدائي يمكن استخدام اختبارات الذكاء العام، حيث إن الاستعدادات الخاصة تتفتح وتتمايز كلما ازداد نضج الطفل وخبرته.

وإذا استُخدمت بطارية اختبارات الاستعدادات الخاصة للتنبؤ بالكفايات المتعلقة بعمل معين يمكن مقارنة الصفحة النفسية للفرد بالصفحة البيانية العامة المتعلقة بمتطلبات العمل لمعرفة مدى التقارب بين الصفحتين. ومن أمثلة هذه البطاريات في

المجالات الأكاديمية والمهنية بطارية الاستعدادات الفارقة Differential Aptitude Tests (DAT) وبطاريات الاستعدادات الحسية الحركية Sensory - Motor Batteries، وفي مجالات انتقاء الطلاب للقبول بالكليات أو الدراسات العليا أو التخصصات المهنية نجد أن هناك عددًا من البطاريات تتمثل في اختبار الاستعداد الأكاديمي Scholastic Aptitude Test (SAT)، واختبار سجل الدراسات العليا Graduate Record Examination (GRE) وغيرهما.

أما البطاريات العامة فتشتمل على اختبارات يمكن استخدامها في مدى واسع من الاستعدادات والمهن، ويمكن اختيار واحد أو أكثر من اختبارات البطارية لاستخدامه في أغراض مناسبة. كما أن هناك اتجاهًا نحو تضمين هذه البطاريات اختبارات لفظية، وتوسيع محتواها بحيث يمكن أن تقيس القدرة العقلية العامة وبعض الاستعدادات الخاصة، وهذا يفيد في زيادة فاعلية التطبيق والحصول على معايير قابلة للمقارنة في جميع اختبارات البطارية.

وهذه البطاريات العامة المتعددة الأغراض أكثر استخدامًا في أغراض الإرشاد التربوي والمهني، ومن أمثلتها بطارية اختبارات الاستعدادات العامة General Aptitude Tests Battery (GATB)، والبطارية المسحية للاستعدادات Employee Aptitude Survey (EAS)، وبطارية فلاناجان لتصنيف الاستعدادات Flanagan Aptitude Classification Tests (FACT).

وتتميز بطاريات الاستعدادات الخاصة بأنه يمكن تطبيقها على مجموعة كبيرة من الأفراد في آن واحد. غير أن هذه البطاريات تقتصر عادة على الورقة والقلم حيث لا تستخدم أجهزة تتطلب أداء عمليًّا معينًا.

اختبارات محاكاة:

وتشتمل على الأنشطة الضرورية المتعلقة بعمل معين إما بتضمين الاختبار مواقف مصغرة لأنماط العمل الفعلي، أو بمحاكاة العمل دون تقديم مواقف فعلية.

وتستند هذه الاختبارات إلى فكرة أن العمل ينبغي أن يُؤدَّى كوحدة متكاملة ترتبط مكوناتها فيما بينها، وهذا الارتباط بين المكونات لا يقل أهمية عن المكونات ذاتها. لذلك فإن مثل هذه الاختبارات تكون أكثر صدقًا في التنبؤ بالنجاح في العمل. ومن ميزات اختبارات المحاكاة Simulation Tests أنها لا تتطلب تحديد القدرات التي ينطوي عليها العمل، وإنما تعتمد على محاكاة العمل ولو بصورة جزئية، وبخاصة في المجالات الصناعية. ومن أمثلة ذلك اختبار محاكاة الاستعداد المتعلق بتشغيل المخارط

Miniature Test For Engine Lathe Aptitude ، واختبار محاكـاة مكبس تثقيب المعادن Miniature Punch Press Test. ويتطلـب بناء مثل هـذه الاختبـارات فكرًا مُبدعًا وقدرة عالية على التصور.

اختبارات موقفية :

تستخـدم هذه الاختبارات فى تـقييم أداء الفرد المـتعلق بعمل مـعين فى ظروف اختبارية مقننة Situational Assessment. وهنا يُطلب من الفرد أداء جـميع العمليات المتعـلقة بعمل معـين أو بعضها فى ظـروف مماثلة تمامًـا لظروف العمل الفعـلية التى سيواجهها الفرد عندما يلتحق بالعمل.

فمثلاً فى مجال إصلاح الأجهزة الكهربائية يمكن أن يقدم للفرد المُختبر جهاز به عطل معين ويطلب منه كشف العطـل وإصلاحه. وفى الأعمال الإدارية أو الكتابية، أو الإرشادية يُقـدم للفرد المختبـر موادٌ متنوعة تُزوده بخـلفية من المعلومـات، ثم يعطى أوراق ومعلومات تماثل ما يتعامل مـعه الإدارى أو المرشد التربوى أو النفسى أثناء أداء كل منهم لمهامه الفعلية، وتسمى هذه المعلومات «معلومات فى السلة In - Basket».

ويتم تقدير درجـات هذه الاختبارات بفحص الاستجابات لتحديـد أنماط اتخاذ القرارات والأحكام المهنية فيهـا. والاختبارات الموقفية Situational Tests لا تقتصر على تقييم الأفراد فقط، وإنما تفيد فى أغراض الـتدريب، فتقدير الدرجات يعتمد أيضًا على معدل ومقدار التحسن فى أداء الفرد المُختبر إذا أُعطى قدرًا من التدرب.

لذلك يمكن أن يُعـاد تطبيق هذه الاختبـارات بعد التحاق الفرد بالـعمل لإعطائه فرصة التحسن، وملاحظة مدى تكامل سلوكه الشخصى والاجتماعى، وكفايته المستقبلية فى تعلم وأداء المـهام التى يقيسهـا الاختبار. وتُستخـدم اختبارات المحاكـاة واختبارات مواقف العمل Work -Sample Tests فى انتقاء العاملين بالجيش والمصانع وتدريبهم، ولكنها ربما تكون قليلة الفائدة فـى المدرسة أو فى أغراض الإرشاد التربـوى.

بعض الافتراضات الأساسية فى قياس الاستعدادات الخاصة :

(١) سبق أن أوضحنا أن الاستعدادات Aptitudes يمكن أن تضم مجموعة من القدرات الخاصة المترابطة التى يُعبَّر عنـها باستجابات الفرد لمواقف سلوكية معينة.

فمـثلاً إذا قلنا أن لدى فـرد استعـدادًا موسيـقيًا، فإن هـذا يعنى أنه يمـتلك مجموعة مـترابطة من المهارات والقدرات والخصائص الـتى تمكنه من تعلم



الموسيقى. أى أن الاستعداد يُعد تكوينًا فرضيًا يُلخص المهارات والقدرات والخصائص المتعلقة به والتى يمكن أن يمتلكها الفرد. لذلك فإن الاستعداد لا يقاس بطريقة مباشرة، وإنما يُستدل عليه من مكوناته السلوكية التى أشرنا إليها، والتى يهدف اختبار الاستعداد لقياسها.

(٢) يُعد قياس الاستعدادات الخاصة نوعًا من القياس مرجعى الجماعة أو المعيار Norm - Referenced ، فدرجة الفرد فى اختبار الاستعداد تقارن بمعيار جماعة مرجعية ينتمى إليها الفرد. فإذا قلنا أن الاستعداد الكتابى لفرد ما مرتفع، فإن هذا يعنى أنه يمكنه تعلم المهارات الكتابية المتعلقة بالملفات والتقارير والمراسلات والطباعة والتدقيق والاختزال أسرع من أقرانه وبجهد أقل. أما إذا كان الاستعداد الكتابى لديه منخفضًا، فإن هذا لا يعنى أنه لا يستطيع تعلم هذه المهارات، وإنما يحتاج لجهد ووقت أكبر مما يحتاجه معظم أقرانه.

(٣) يفترض فى قياس السمات الإنسانية أنها مستقرة نسبيًا ، ونظرًا لأن الاستعدادات تفيد فى التنبؤ المستقبلى بأداء الأفراد، فإنها ينبغى أيضًا أن تتسم بالاستقرار النسبى عبر الزمن. ولكن هذا لا يعنى أن الاستعدادات تُعد سمات موروثة لا تتغير أو تتحسن طوال حياة الفرد، فالاستعدادات لها أساس وراثى بلا شك، ولكنها تعكس أيضًا خبرات الفرد وتعلمه، أى أنها نتاج تفاعل العوامل الوراثية والعوامل البيئية كما ذكرنا فى مستهل هذا الفصل. والاستقرار النسبى لدرجات اختبارات الاستعدادات يعكس اتساق نمط حياة معظم الأفراد.

(٤) إن الاستعداد المرتفع لدى فرد فى مجال معين لا يعنى أنه سوف يُنمّى قدرته فى هذا المجال دون تدريب مناسب. أى أن التنبؤ استنادًا إلى أحد اختبارات الاستعدادات يعنى أن تجمّع من المهارات والقدرات الخاصة التى تُمثل هذا الاستعداد مع تدريب مناسب سوف يؤدى إلى سلوك معين.

ونظرًا لأن التدريب الذى يُعد مناسبًا لفرد ما قد لا يناسب فردًا آخرًا، فإن أساليب التدريب ينبغى أن تتباين بتباين الاستعدادات. لذلك فإن كرونباك Cronbach وسنو Snow (١٩٧٧) أكدا أهمية التفاعل بين الاستعدادات وأساليب التدريب أى المعالجات Aptitude - Treatment Interaction (ATI) لكى يكون أسلوب التدريب فاعلاً مما يزيد من دافعية الفرد ويؤدى إلى تحسن أدائه.

ولعل هذا ما دعا بعض علماء النفس والقياس النفسى إلى إعادة النظر فى مفهوم الاستعداد Aptitude الذى أوضحناه فيما سبق، حيث يرى بلوك Block ، وبلوم Bloom، وكارول Carroll وغيرهم أن أسلوب التدريب الجيد والمناسب لاستعدادات الفرد واهتماماته ويستثير دافعيته للتعلم يؤدى إلى تقليل الزمن الذى يحتاجه الفرد فى إتقان المهارات المرجوة. أما إذا كان أسلوب التدريب غير مناسب للفرد، وبخاصة إذا كان استعداده منخفضًا فسوف يحتاج إلى زمن أطول لتعلم المهارات المعينة.

لذلك رأى هؤلاء العلماء أن الاستعداد يمكن تعريفه من حيث الزمن الذى يحتاجه الفرد لتحقيق هدف معين أو التمكن من مهارة. فإذا تباينت المعالجة التعليمية أو التدريبية بتباين استعدادات الأفراد بحيث تناسب المعالجات استعدادات الأفراد، وأتيح لكل منهم الوقت الذى يحتاجه للتعلم، فإن معظمهم يستطيع إتقان المهارة أو المطلب المرجو تحقيقه.

أى أن الاستعداد يتمثل فى الزمن الذى يحتاجه الفرد للتعلم ولا يتمثل فى مستوى صعوبة أو تعقد المادة المتعلمة. ولعل ذلك يرتبط بالقياس مرجعى المحك بدرجة أفضل من ارتباطه بالقياس مرجعى الجماعة أو المعيار. فمحك الأداء Performance Criterion هو الإطار المرجعى المناسب لقياس الاستعدادات، حيث إنه أكثر ملاءمة من معايير الجماعة Group Norms.

أساليب بناء اختبارات الاستعدادات الخاصة :

تختلف أساليب بناء اختبارات الاستعدادات الخاصة باختلاف نوع الاختبار وتكوينه والغرض الذى يُستخدم من أجله، وسوف نلقى الضوء فيما يلى على بعض هذه الأساليب:

أساليب تستند إلى التحليل النظرى لمكونات الاستعداد :

فعند بناء مقياس للاستعداد الموسيقى أو الفنى مثلاً، يمكن تحديد الخصائص أو القدرات الأساسية التى تسهم فى الأداء الفاعل فى هذه المجالات. فالقدرات الخاصة المتعلقة بالاستعداد الموسيقى تتعلق بتمييز النغم، والإيقاع، وشدة الصوت وغير ذلك من المكونات التى يمكن أن يقترحها المتمرسون فى المجالات الموسيقية والتى تتأثر بنوعية التدريب.

أساليب إمبيريقية :

فمعظم اختبارات الاستعدادات الخاصة تهدف للتنبؤ بمحك أداء مستقبلى، لذلك يُراعى فى بناء هذه الاختبارات أن ترتبط درجاتها بمحك مناسب لكى يكون الصدق

التنبؤى Predictive Validiy مرتفعًا، حيث إن هذا يعد من الشروط الأساسية فى هذا النوع من الاختبارات. فإذا اشتمل الاختبار على عدة اختبارات فرعية، فإنه ينبغى أن يكون الارتباط بين درجات هذه الاختبارات منخفضًا أو صفريًا، بينما يكون ارتباطها بدرجات المحك مرتفعًا، وبذلك يتم انتقاء المفردات على أساس إمبيريقى. ويمكن الرجوع إلى الفصل الخامس لمراجعة تفاصيل ذلك.

أساليب تستند إلى التحليل العاملى:

تعتمد كثير من بطاريات الاستعدادات الخاصة فى بنائها وتكوينها على نتائج التحليل العاملى Factor Analysis، حيث يكون الهدف هو بناء بطاريات يشتمل كل منها على عدد محدود من الاختبارات الفرعية التى يمكن باستخدامها التنبؤ بدرجة أفضل بالنجاح فى تخصصات دراسية معينة أو مهن مختلفة.

ونظرًا لأن طبيعة المهن ومكوناتها ومواصفاتها دائمة التغير، فإن هذا يتطلب المراجعة المستمرة للاختبارات التى تشتمل عليها البطاريات استنادًا إلى نتائج التطبيق الميدانى المستمر لهذه البطاريات وما يُستمد من بيانات متجددة.

أساليب تستند إلى نتائج تحليل العمل:

تُصمم عادة اختبارات الاستعدادات الخاصة لقياس مهارات محددة ومعارف تتعلق بمجال مهنى معين. فاختبارات الاستعداد الميكانيكى أو الكتابى أو الفنى مثلاً تقيس المكونات السلوكية التى تتضمنها هذه الاستعدادات.

لذلك فإن بناء هذه الاختبارات يتطلب تحديدًا دقيقًا لما تشتمل عليه هذه المكونات من مهارات ومعارف استنادًا إلى تحليل العمل Job Analysis. والهدف من هذا التحليل تحديد الأداء الفاعل فى مهنة أو مجال معين. ونظرًا لأن كثيرًا من المؤسسات الصناعية والتقنية وغيرها من المؤسسات المهنية تود اتخاذ قرارات انتقاء الأفراد الفنيين العاملين بها على أسس علمية، فإنها تقوم ببناء اختبارات استعدادات مناسبة للأعمال أو المهن المتعلقة بها، وبخاصة إذا لم تتوافر لديها اختبارات جاهزة.

ففى هذه الحالة ينبغى أن يستند بناء اختبارات الاستعدادات على نتائج تحليل العمل المتعلق بهذه المهن، مثل أعمال النجارة، والكهرباء، والخراطة، وإصلاح الآلات، وترويج السلع، وما إلى ذلك.

إجراءات تحليل العمل :

يستند تحليل العمل Job Analysis على إجراءات منهجية منظمة نوضحها فيما يلي:

(١) تعريف المقصود بالعمل :

فالعمل Job هو تجمُّع متجانس نسبيًا من المهام الوظيفية التي يقوم بها الفرد لتحقيق غرض أساسي في مؤسسة معينة. والتجانس النسبي لهذه المهام قد يعني تماثلها في النوع أو المحتوى أو المتطلبات المشتركة، أو قد يعني أن الفرد الذي يستطيع أداء إحدى هذه المهام ربما يستطيع أداء مهمة أخرى. فتعريف المقصود بالعمل في ضوء المهام Work Tasks المتعلقة به يساعد في تركيز الاهتمام على السلوك الفعلي المتعلق بالعمل ومتطلباته، وهذا هو الهدف الأساسي من تحليل العمل. غير أنه ينبغي مراعاة احتمال تغير متطلبات العمل ومواصفاته بتغير مهارات الأفراد، أو طبيعة العمل والمشرفين عليه، أو بيئة العمل الفيزيائية والنفسية. أي أن ديناميكيات العمل يجب أن تؤخذ بعين الاعتبار عند تحديد المقصود بالعمل.

(٢) تحديد طبيعة العمل ومواصفاته :

إن تحليل العمل يصف الجوانب المهمة لعمل معين التي تميزه عن غيره من الأعمال. لذلك فإن تحديد طبيعة العمل ومواصفاته يزودنا بوصف ما يفعله الفرد في العمل لكي يتسنى بناء أدوات انتقاء الأفراد وتصنيفهم في ضوئه.

ويجب أن يتضمن هذا الوصف العمليات التي يؤديها الفرد، والأجهزة المستخدمة، وظروف العمل ومخاطره، والتدريب الذي يتطلبه، وعلاقة العمل بغيره من الأعمال. وهذا الوصف أساسي حيث تُحدد في ضوئه مواصفات العمل Job Specifications ، أي قائمة الكفايات التي يستند إليها بناء أدوات القياس؛ لذلك يجب أن تكون هذه القائمة مرتبطة بالعمل المحدد ومتطلباته التي تميزه عن غيره من الأعمال. كما يجب أن تُحدد القائمة جوانب الأداء التي تُميز تمييزًا دقيقًا بين الأداء الجيد والأداء الضعيف.

فالأفراد من مستويات كفاية متباينة ربما يختلفون في طريقة أدائهم معظم جوانب العمل، لذلك فإن هناك جوانب معينة من العمل هي التي توضح الفروق بين المتميزين والضعفاء، ويطلق على هذه الجوانب التمييزية «الأنشطة أو المتطلبات الحرجة Critical Incidents». وهذه الجوانب لا تفيد فقط في بناء اختبارات الانتقاء، وإنما أيضًا في تصميم برامج التدريب أو تعديل البرامج القائمة، وتطوير إجراءات العمل، وتقليل مخاطره، وتطوير تصميم الآلات والأجهزة في المؤسسات الصناعية.

(٣) أساليب جمع المعلومات المتعلقة بالعمل :

إن تحديد المواصفات التفصيلية للمهام في الخطوة السابقة يتطلب جمع معلومات تتعلق بالعمل من مصادر متعددة بأساليب منظمة لكي يتسنى الإفادة منها. إذ يمكن الاطلاع على المطبوعات ذات العلاقة، مثل أدلة التشغيل والتدريب، والكتيبات الخاصة بأعمال مماثلة، وسجلات الأداء، وبخاصة تلك التي تتضمن مواصفات الأخطاء الشائعة في الأداء، وصعوبات التعلم، وأسباب إخفاق بعض الأفراد في العمل، كما يمكن فحص سجلات شكاوى العملاء، وبخاصة في الأعمال التجارية الترويجية. وكذلك يمكن ملاحظة عينات من أداء الأفراد العاملين، ولكن ينبغي مراعاة عدم تأثر الأداء بوجود القائم بالملاحظة. وللتغلب على ذلك يمكن استخدام أسلوب الملاحظة بالمشاركة Participant Observation ،حيث يقوم الفرد الملاحظ بالمشاركة في العمل والوظائف كما لو كان واحداً من العاملين، ويسجل ملاحظاته عنهم أثناء مشاركاته معهم.

كذلك تُعد المقابلة الشخصية مصدراً مهماً من مصادر جمع المعلومات المتعلقة بالأعمال المختلفة. ويمكن أن يُجرى المقابلة المشرفون على العمل، وأخصائيو التدريب والعاملين القدامى من ذوى الخبرة والأداء المتميز. غير أن التقارير التى يقدمها القائمون بالمقابلة قد تتسم أحياناً بالذاتية أو تتضمن تعبيرات أو مصطلحات غير متفق عليها، لذلك ينبغي أن تركز التقارير على الحقائق الموضوعية وما يؤديه الفرد العامل بالفعل.

ويمكن جعل هذه التقارير أكثر موضوعية باستخدام مذكرات قيد الأعمال أو الأنشطة Activity Logs بدلاً من الاعتماد على ذاكرة القائم بالمقابلة. وكذلك يمكن استخدام «أسلوب الأنشطة الحرجة (CIT) Critical Incidents Technique» الذى اقترحه فلاناجان (Flanagan, 1954) والذى أشرنا إليه في الخطوة السابقة. ويتم إجراء هذا الأسلوب في مدة زمنية محددة، على أن يتضمن التقرير الأنشطة أو السلوك كما حدث، مع ضمان سرية البيانات التى يتضمنها التقرير.

ومن الأساليب الأخرى المهمة التى يمكن عن طريقها الحصول على معلومات منظمة وبيانات كمية تتعلق بتحليل العمل الاستبيانات أو القوائم Inventories التى يتم إعدادها مسبقاً، حيث تدون فيها المهام الأساسية المتعلقة بالعمل والتى أمكن الحصول عليها باستخدام بعض الأساليب السابقة، وكذلك فقرات تتعلق بظروف العمل وغير ذلك من الخصائص.

وتوجد أنواع متعددة من هذه الاستبيانات بعضها يتضمن ميزان تقدير رقمى متدرج Rating Scale يتطلب من الفرد وضع علامة على رقم التقدير الذى ينطبق على

ما يؤديه من عمل، أو على الزمن الذى يستغرقه في مهـمة معينة بالنسبة للمهام الأخرى المتعلقة بعمل مـعين، وبعضها الآخر يتضمن مـيزانًا بيانيًا خطيًا يُحدِّد عليـه الفرد نوعية الأداء الذى يتراوح بين أداء متميز وأداء ضعيف، ويضع علامة على الخط تناظر المستوى الذى يراه مناسبًا له.

وهذان النوعان من الاستبيـانات يتعلقان بمهام العمل Job - Oriented ، لذلك فإن استخدامهما يقتصر عـلى مجال مهنى معين صممت من أجله. غير أن هناك أنواعًا أخرى من الاستبيانات تتعلق بالفرد العامل نفسه، وخصائصه النفسية التى تسهم فى رفع مستوى الأداء فى العمل وتتناول سلوك الفـرد فى العمل الذى يتمـيز بدرجة أكـبر من العمومـية، ومثال ذلك استبـيان تحليل العمل Position Analysis Questioinnaire الذى أعده ماكورميك McCormic، وجينرت Jeanneret، وميكام Mecham. ويشتمل هذا الاستبيان على ١٩٤ عنصراً متعلقًا بالفرد العامل وأدائه يمكن تصنيفها فى عدة أقسام مثل: مصـادر حصول العامل علـى معلوماته المتعلقـة بالعمل، والعمليات العـقلية التى يُوظفها فـى عمله كالاستدلال واتخـاذ القرار والتخطيط، والأنشطة البـدنية، والمهارات اليدوية والأدوات والأجهزة التى يستخدمها، وعلاقة العامل بزملائه ورؤسائه فى العمل، وبيئة العمل، وغير ذلك من الخصائص.

أى أن هذه الاسـتبـيانات تركـز على مُدخـلات المعلومـات لدى الفرد العـامل، وتجهيزه ومعـالجته للمعلومات، والفعل أو الاستجـابة الناتجة. وتتطلب من الفرد إما أن يضع علامة على العنصر الذى ينطبق عليه، أو يحدد على ميزان التقدير الرقم الذى يدل على مدى الأهمية، أو الزمن أو الصعوبة فيما يتعلق بكل عنصر فى هذه الأقسام.

وقد استُـخدم هذا الاستبيـان فى تحليل ٥٣٦ نوعًا مختلفًا من الأعـمال، وأُجرى تحليل عاملى للدرجـات الناتجة، حيث أمكن اسـتخلاص خمـسة عوامل تنطوى عليـها استجـابات الأفراد العاملين فى الأنواع المختلفة من الأعـمال تتلخص فى : القدرة على اتخاذ القـرارات والتواصل، والمسئـولية الاجتـماعية، والقـدرة على أداء أنشطة تتطلب مهارات، والنشاط الحـركى البدنى فى بيئة عمل مناسبة، والقدرة على تـشغيل الأجهزة التى تتطلب عمـليات حسيـة وإدراكية ووظائف جـسمانيـة، وعمليات تجـهيز ومعـالجة المعلومات.

أى أن هذا النوع من الاستبيـانات يساعد فى الحصول على قيـاس كمى لكثير من المتغيرات النفسية والبدنية والاجتمـاعية والمعرفية التى تسهم فى أداء المهام المتعلقة بالأنواع المختلفـة من الأعمـال، وكذلك ظروف العمل وخـصائصـه الفيزيائيـة التى تتطلب من الأفراد العاملين التكيف معها.

(٤) انتقاء أو بناء أدوات القياس :

بعد التحديد المبدئي للمهارات والمعارف والقدرات الخاصة التي تبدو أهميتها في أداء نوع معين من الأعمال، فإن الخطوة التالية تتعلق بترجمتها إلى إجراءات ومواقف اختبارية.

ويمكن الاعتماد في ذلك على الاختبارات والمقاييس المتوافرة التي يتبين ملاءمتها لقياس مهارات أو قدرات مشابهة لما أسفرت عنه نتائج تحليل العمل. أما إذا لم تكن هناك اختبارات أو مقاييس مناسبة، فإنه ينبغي ابتكار مهام اختبارية لقياس هذه القدرات والمعارف والمهارات المتعلقة بعمل معين. وهذا يتطلب اتباع جميع الخطوات التي سبق أن أوضحناها في الفصل الثامن عند مناقشتنا لكيفية بناء الاختبارات المقننة. ويفضل تجريب الاختبار على عينة استطلاعية مناسبة للتحقق من وضوح مفرداته وتعليماته، وتحديد زمن إجابته إذا كان من نوع الاختبارات الموقوتة، وتحليل مفرداته والتحقق من ثبات درجاته. أما فيما يتعلق بصدق الاختبار وما يتعلق بذلك من اختيار المحك المناسب والتحقق من صدقه، وتحديد درجة القطع وما يترتب عليها من أخطاء اتخاذ قرارات الانتقاء أو التصنيف، فقد أوضحناه بالتفصيل في الفصل الخامس.

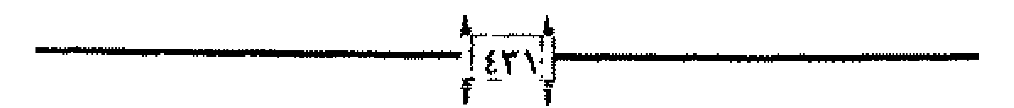
ومن الجدير بالذكر أنه ليس من الضروري الاعتماد على الاختبارات التي تتطلب ورقة وقلم، وإنما يمكن استخدام اختبارات محاكاة، أو اختبارات تتضمن عينات من السلوك أو الأداء الفعلي في العمل كما أوضحنا في هذا الفصل. غير أنه في جميع الحالات ينبغي إجراء دراسات للتحقق من الصدق التنبؤي لهذه الأدوات.

اختبارات الاستعدادات الخاصة :

ذكرنا فيما سبق أنه توجد أنواع متعددة من اختبارات الاستعدادات الخاصة في المجالات الأكاديمية والمهنية أوضحناها بالشكل التخطيطي (١٠ ـ ١) .

ففي أغراض التوجيه والإرشاد المهني والأكاديمي، وانتقاء الأفراد تبدو أهمية قياس قدرات ومهارات ومعارف الأفراد في هذه المجالات، مما قد يتطلب تطبيق عدة اختبارات مختلفة. غير أن تطبيق هذه الاختبارات المنفصلة يستغرق وقتًا طويلاً، كما أن معاييرها تكون مستمدة من جماعات مرجعية مختلفة، وبذلك لا نستطيع المقارنة بين درجات الفرد في أحد هذه الاختبارات بدرجاته في اختبار آخر مما يجعلها محدودة الفائدة في أغراض الإرشاد.

لذلك يُفضل في هذه الحالة استخدام بطاريات الاستعدادات المتعددة Multiple Aptitude Batteries، حيث يمكن تطبيقها على مجموعة كبيرة من الأفراد في آن



واحد، ورسم صفحة نفسية (بروفيل) للفرد بـاستخدام درجات الاختبارات الفرعية التى تشتمل عليها البطارية.

وسوف نعرض فيمـا يلي بعض أمثلة من هذه البطاريات الفارقة والـعامة الشائعة الاستخدام فى المجالات الأكاديمية والمهنـية، ثم نعرض أمثلة أخرى لبعض اختبارات الاستعدادات المتعلقة بمجالات نوعية محددة.

بطاريات الاستعدادات المتعددة :

(١) بطارية اختبارات الاستعدادات الفارقة

Differential Aptitude Test Battery (DAT):

تعد من أكثر بطاريات الاستعدادات استخدامًا وتميزًا، وقد أعدها بنيت Bennet، وسيشور Seashore، ووزمـان Wesman، وصـدرت طبعـتـها الأولى عـام ١٩٤٧، وأجريت عليها بعد ذلك عدة تعديلات فى طبعاتهـا التالية حتى صدرت الطبعة الخامسة عام ١٩٧٤. ويتفق كـثير من علماء النفس والتربية على أن هذه البطارية تتميز بجودة بنائها وطريقـة عرضها لسلسلة الاختبـارات التى تشتمل عليها، حيـث استند بناؤها إلى إطار نظرى وعملى متوازن، وروعى أن يكون الارتباط بين اختباراتها الفرعية منخفضًا، وأن تقيس مـجموعة من القدرات الـخاصة التى يمكن أن تـساعد فى التوجيـه التربوى والمهنى لطلاب المرحلة الثانوية.

وعلى الرغم من أن مصممى البطارية لم يعتـمدوا فى بنائها على أسلوب التحليل العاملى، إلا أنهم استرشدوا فى انتقائهم للاختـبارات التى تشتمل عليها بنتائج دراسات التحليل العـاملى للاختبارات المتوافرة، وكـذلك بالاحتياجات الفعلية لـعملية الإرشاد التربوى والـمهنى، أى أنهم حاولـوا انتقاء اختـبارات يمكن الإفادة مـنها عمليًـا وتتسم بالثبات والصدق.

كما تتميز البطارية بسهولة التصحيح، وتعتمد على القوة. ولكن تطبيقها يستغرق ما يقرب من ثلاث ساعـات، ويتراوح زمن تطبيق أىٍّ من اختباراتهـا الفرعية من ٦ إلى ٣٠ دقيقـة ، ويمكن تطبـيق أىٍّ من هذه الاختبـارات على حدة. وتتكون الـبطارية فى طبعتها الخـامسة من صيغتين متكافئـتين ، واستمدت معاييرها فى البيـئة الأمريكية من عينة قومـية اشتملت على ٦٤٠٠٠ طالب فى مـرحلة التعليم الثانوى. وأُعـدت معايير مستقلة لـكل من البنين والبنات، ويمكن رسم صفـحة نفسية (بروفيل) للرتب الـمئينية المقابلة للدرجـات الخام لكل طـالب بحيث يمكن أن يستند إليها الـمرشد التربوى فى تعريف الفرد بمستوى استعداداته المتعددة لكى يتخذ الفـرد قراره فى ضوئها. ويمكن الرجوع إلى شكل (٦ ـ ٤) فى الفصل السادس للاطلاع على مثال لهذا البروفيل.



وتشتمل البطارية على ثمانية اختبارات فرعية هي : الاستدلال اللفظي، (المتشابهات)، والقدرة العددية (العمليات الحسابية)، والاستدلال المجرد، والعلاقات المكانية، والاستدلال الميكانيكي، والسرعة ، والدقة الكتابية، والهجاء، واستخدام اللغة (النحو ، والترقيم).

وبذلك يكون لكل فرد تسع درجات بما في ذلك الدرجة الكلية، ويمكن أيضًا الحصول على مجموع درجتي كل من اختباري الاستدلال اللفظي، والقدرة العددية كمؤشر للاستعداد الدراسي العام. وقد وُجد أن هذا المؤشر يرتبط باختبارات القدرة العقلية العامة ارتباطًا تزيد قيمته عن ٠,٧٠ ، غير أنه لوحظ أن قيم معامل الارتباط بين الاختبارات الفرعية التي تشتمل عليها البطارية تراوحت بين ٠,٠٦ ، ٠,٦٧ ، مما يدل على أن هذه الاختبارات لا تقيس «عوامل» مستقلة، وبذلك تحتاج البطارية لمزيد من الدراسات للتحقق من أنها بطارية فارقة بالفعل، أي التحقق من قدرتها التمييزية. وقد أجريت دراسات متعددة حول القيمة التنبؤية للبطارية ، وأوضحت الارتباطات بين درجات اختباراتها الفرعية وكل من التقدير العام للطلاب في المقررات المختلفة، واختبارات الذكاء، ودرجات بطاريات استعدادات أخرى، واختبارات تحصيل موضوعية، وبيانات تتعلق بالمهن، أن هناك علاقة فارقة بين درجات اختبارات البطارية والتحصيل في المجالات الدراسية المختلفة، ومستويات القدرة العقلية، والتصنيفات المهنية، وتقديرات المشرفين على العمل في مجالات مهنية محدودة.

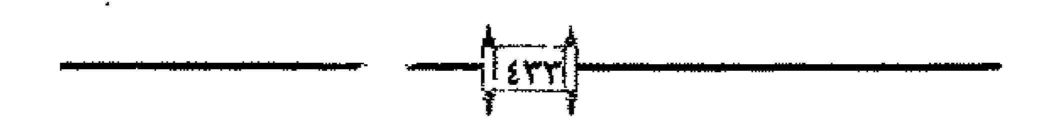
وقد تراوحت قيم معامل ثبات درجات الاختبارات الفرعية باستخدام طريقة التجزئة النصفية بين ٠,٨٧ ، ٠,٩٧ للبنين، وبين ٠,٨٣ ، ٠,٩٦ للبنات في الفرق المختلفة.

(٢) بطارية اختبارات الاستعدادات العامة

: General Aptitude Test Battery (GATB)

أعد هذه البطارية مكتب خدمات التوظيف بالولايات المتحدة الأمريكية، وتختلف عن البطارية السابقة في أنها تُستخدم في أغراض الانتقاء والتسكين في المجالات الصناعية والتجارية بخاصة. فبطارية الاستعدادات الفارقة أجرى تقنينها في الإطار المدرسي، وتستخدم أساسًا في التنبؤ بالنجاح الأكاديمي.

وقد صُممت بطارية اختبارات الاستعدادات العامة استنادًا إلى تحليل العمل، والتحليل العاملي لمجموعة من الاختبارات بلغ عددها ٥٩ اختبارًا متعلقًا بأعمال مهنية مختلفة، حيث تبيّن أن هذه الاختبارات تنطوي على عوامل أو قدرات مختلفة لا يزيد عددها عن عشرة، لذلك أعدت هذه البطارية لقياس معظم هذه العوامل.



فالطبعة الحديثة لهذه البطارية اشتملت على ١٢ اختبارًا فرعيا يمكن باستخدامها الحصول على درجات لتسعة عوامل أو قدرات مختلفة كالتالي:

القدرة العقلية العامة (وتتضمن القدرة اللفظية، والقدرة العددية ، والقدرة المكانية) ، وإدراك الصيغ ، والإدراك الكتابى، والتآزر الحركى، ومهارة الأصابع، والمهارة اليدوية.

وتُقاس هذه العوامل أو القدرات بثمانية اختبارات ورقة وقلم، وأربعة اختبارات تعتمد على أجهزة. وبعض اختبارات الورقة والقلم تشبه اختبارات بطارية الاستعدادات الفارقة.

وهذه البطارية تناسب طلاب المرحلة الثانوية والراشدين. وتُحوَّل الدرجات الخام فى العوامل التسعة إلى رتب مئينية أو درجات مُحوَّلة متوسطها ١٠٠ وانحرافها المعيارى ٢٠، وبذلك يمكن مقارنة الدرجات المُحوَّلة للفرد بأنماط الاستعدادات المهنية التى يبلغ عددها حوالى ٣٦ نمطًا استُمدت من تحليل درجات البطارية من عينة من الأفراد العاملين فى أكثر من ٨٠٠ مهنة.

ولتيسير استخدام أنماط درجات الاختبارات التى تشتمل عليها البطارية فى أغراض الإرشاد المهنى، تم تجميع المهن التى تتطلب استعدادات متماثلة فى عدد قليل من المجموعات المهنية. وحُددت درجات قطع للاستعدادات الثلاثة الأكثر أهمية لكل مجموعة مهنية بحيث تُمثل الحد الأدنى للمتطلبات التى تُسهم فى الأداء الفاعل فى كل من هذه المجموعات.

ومما يؤخذ على هذه البطارية اعتماد الاختبارات على عامل السرعة فى الإجابة. وعدد الاستعدادات التى تقيسها محدود، حيث إنها لا تتضمن اختبارات تقيس الفهم الميكانيكى، أو الاستدلال ، أو الابتكار، وكذلك فإن الاستناد إلى درجات القطع المتعددة فى الاختبارات فى عمليات الانتقاء والتسكين بدلاً من استخدام معادلات الانحدار يحتاج إلى مزيد من الدراسات.

(٣) بطارية اختبارات سجل الدراسات العليا

Graduate Record Examination Aptitude Test (GRE):

تستخدم هذه البطارية من الاختبارات لانتقاء الطلاب الذين يودون الالتحاق بالدراسات العليا بالكليات المختلفة فى الولايات المتحدة الأمريكية، وتطبق فى مختلف دول العالم لانتقاء طلاب البعثات والمنح الدراسية.

وتنقسم البطارية إلى قسمين رئيسيين: أحدهما قسم عام يشتمل على اختبارات تقيس الاستعداد اللفظى، والاستعداد الكمى، والاستدلال، وقسم متقدم يقيس

التحصيل فى حوالى ٢٠ تخصصًا أكاديميًا رئيسًا، مثل علم النفس، والتاريخ، والكيمياء وغيرها.

واختبارات الاستعداد اللفظى تتضمن التعرف على المتضادات، واستخدام المتشابهات، وفهم الفقرات، وتتضمن اختبارات الاستعداد الكمى، الاستدلال الحسابى، والجبر، والهندسة.

وتُحوَّل الدرجات الخام فى البطارية إلى درجات محوَّلة متوسطها ٥٠٠، وانحرافها المعيارى ١٠٠. وتتطلب بعض الكليات حدًا أدنى للدرجات الكلية المحوَّلة للقبول. وقد أُعدت مطبوعات وكتيبات وأدلة ودورات تدريبية تُعين فى التدرب على إجابة اختبارات هذه البطارية لكى تزيد فرص حصول الطالب على درجات مرتفعة، على الرغم من عدم وجود بيانات مُوثقة تُبين فاعلية هذه المطبوعات والتدريبات فى تحقيق الغرض منها.

وقد أوضحت نتائج استخدام معامل التجانس الداخلى اتساق درجات الاختبارات الفرعية بدرجة مقبولة، غير أنه لا توجد أدلة كافية تُبيِّن مدى صدق البطارية فى التنبؤ بنجاح الطلاب بعد التحاقهم بالدراسات العليا بالكليات المختلفة.

اختبارات استعدادات تتعلق بمجالات نوعية محددة :

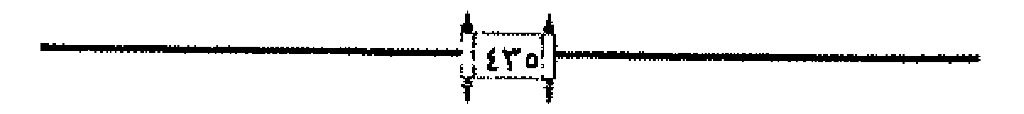
تهدف اختبارات الاستعدادات التى تتعلق بمجالات مهنية أو دراسية معينة التنبؤ باحتمال نجاح الأفراد فى هذه المجالات. ومن بين الاستعدادات الخاصة التى سنوضحها فيما يلى:الاستعداد الميكانيكى، والاستعداد الكتابى، والاستعدادات الموسيقية، والفنية، والابتكارية.

وسوف نُقدم أمثلة لبعض الاختبارات التى تقيس هذه الاستعدادات. ولعل اختيارنا لهذه المجالات يرجع إلى أن كثيرًا من الأفراد يهتمون بالتخصص فى دراستها أو العمل فى مهن متعلقة بها مما يتطلب توافر الاستعدادات المناسبة لديهم، وتهدف برامج الإرشاد والتوجيه التربوى والمهنى للتحقق من ذلك.

الاستعداد الميكانيكى :

تعتمد كثير من المهن التى يشغلها عدد كبير من الأفراد على الاستعداد الميكانيكىMechanical Aptitude، وتتباين هذه المهن فى مستوى المهارة التى ينبغى أن تكون متوافرة لدى الأفراد الذين يعملون فيها.

فأعمال ميكانيكا السيارات والآلات والأجهزة وبناء السفن، وأعمال الخراطة



والسباكة والنجارة تُعد أمثلة لهذه المهن. والتنبؤ بدرجة كبيرة من الثقة بنجاح الأفراد الذين يرغبون في الانتماء إليها يتطلب تطبيق اختبارات متعددة لقياس الاستعداد الميكانيكي لديهم.

ويرى بعض علماء النفس أن هذا الاستعداد يُعد قدرة واحدة تختلف درجتها من فرد إلى آخر، بينما يرى البعض الآخر أنه يشتمل على عدد من القدرات المستقلة نسبيا، غير أن الدراسات العاملية المتعلقة بهذا الاستعداد أكدت تعدد القدرات الخاصة به مما يتطلب قياسه باختبارات متعددة أو ببطارية من الاختبارات المناسبة.

فالمهن المختلفة في هذا المجال تتطلب مهامًا مختلفة مما يدعو إلى قياس القدرات الخاصة المطلوبة للأداء الجيد باستخدام مجموعات مختلفة من الاختبارات المتعلقة بالاستعداد الميكانيكي.

وهذا الاستعداد لا يشير إلى سلوك حركي أو عضلي أو يدوي كما يظن البعض، وإنما يشير إلى وظائف عقلية رمزية، فقد أوضحت الدراسات التي حاولت تحليل الاستعداد الميكانيكي أنه يشتمل على مكونتين رئيستين هما: الفهم الميكانيكي المتعلق بفهم العلاقات الميكانيكية والفيزيائية، والإدراك أو التصور البصري المكاني المتعلق بالتعامل الرمزي مع الأشكال والصيغ في الفراغ.

لذلك نلاحظ أن اختبارات الاستعداد الميكانيكي اعتمدت في بداية تطورها على قياس الأداء المهاري الفعلي حيث يُقدم للفرد المُختبر معدّات ميكانيكية بسيطة مفككة ويطلب منه تجميعها. غير أنه لوحظ أن هذه الاختبارات تتأثر تأثرًا واضحًا بخبرة الفرد، حيث يمكن أن يحصل فرد لديه خبرة بهذه المُعدّات على درجة مرتفعة في الاختبار على الرغم من انخفاض استعداده الميكانيكي. وهذا يدل على أن هذه الاختبارات لا تصلح لقياس هذا الاستعداد، علاوة على أنها تقيس متغيرات تتعلق بالمهارة اليدوية والتآزر البصري الحركي.

لذلك اتجه العلماء إلى قياس المكونتين الرئيسيتين اللتين ذكرناهما أعلاه المتعلقتين بالوظائف العقلية الرمزية.

(١) اختبار الفهم الميكانيكي

Bennett Mechanical Comprehension Test (BMCT) :

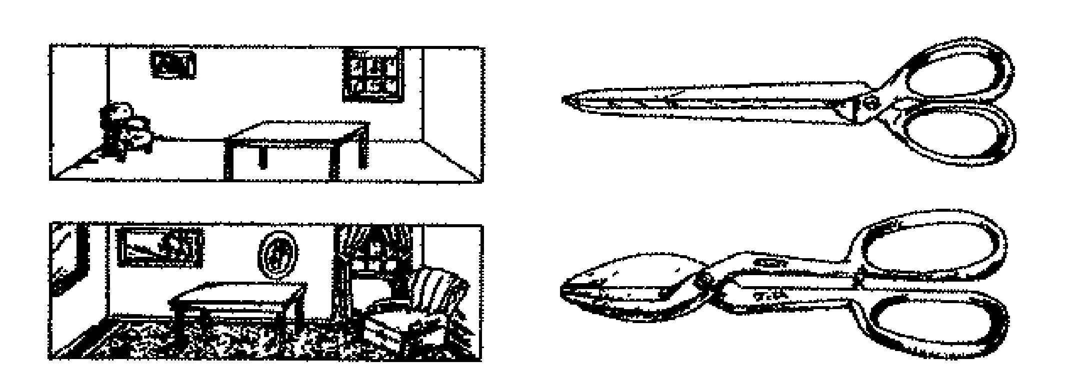
أعد هذا الاختبار بنيت Bennett، وصدرت طبعته المعدلة عام ١٩٧٠، وهو من الاختبارات شائعة الاستخدام لقياس فهم المبادئ الميكانيكية المتعلقة بالأدوات التي تستخدم بكثرة في الحياة اليومية، مثل المفك، والشاكوش، ومفتاح الصمولة، والمقص، وغير ذلك.

ويُعد هـذا الاختبـار مماثلاً أيـضًا لأحد الاختـبارات الفـرعية المـتعلق بـالقدرة الميكانيكية في بطارية الاستعدادات الفارقة (DAT) التي أشرنا إليها فـيما سبق. ويقيس هذا الاختبار فهم العلاقات الميكانيكية، وتطبيقات المبادئ الفيزيائية في مواقف عملية. والاختبار عـبارة عن كتيب يتضـمن ٦٠ رسمًا تخطيطيًا لمواقف ميكانيكيـة لا تحتوى بالضرورة على أدوات يدوية، مثل العلاقات بين الحبال والبكرات، والعجلات، والقوى الفيزيائية، والتركيبات الميكانيكية.

ولا يتطلب الاختـبار أداء عمـليًا وإنما يتـطلب تفكيرًا رمـزيًا، ولا يعتمـد على المعارف العامة أو تذكر الأشياء وإنما يعتمد على الفهم، والتحليل المنطقى، واستيعاب العلاقات بين الأشياء.

فمـثلاً يقدم لـلفرد رسم تخـطيطى لمقـصين ويطـلب منه تحـديد أيهمـا يمكن استخدامه بـدرجة أفضل فى قطع المعادن، أو غرفـتين تحتويان على قطع مـختلفة من الأثاث، ويطلب منه تحديد الغرفة التى يحدث فيها صدى أعلى للصوت.

ويوضح الشكل (١٠ ـ ٢) التالى هاتين المفردتين :



شكل (١٠ ـ ٢) يوضح بعض مفردات اختبار الفهم الميكانيكى

وقد أُعدَّت ثلاث صيغ للاختبار تتفاوت فى درجـة صعوبتها، حيث تُطبق صيغة معينة منها على المتقدمين للعمل فى المـستويات المهنية العليا، ويستغرق تطبيق الاختبار عادة حوالى نصف السـاعـة، ولكنه يعتمد على القوة. وقد صُمِّم الاختـبار لكى يناسب طلاب نهاية مرحلة التعلـيم الثانوى، والمتقدمين للالتحاق بالمهـن الصناعية والعاملين فيها. وتُفسـر درجات الاختبار فى ضوء معايير مئينية خاصة مستـمدة من طلاب نهاية المرحلة الثانوية العامة والصناعية وطلاب السنة الأولى بكليات الهندسة، ومعايير خاصة للمتدربين فى المصانع، والمهندسين وغيرهم من العاملين، ومعايير خاصة بالإناث.

ويشير دليل الاختبار إلى أن قيم معاملات صدق الاختبار فى التنبؤ بالنجاح فى المهن المختلفة التى تتطلب الاستعداد الميكانيكى تتراوح بين ٠,٣٠ ، ٠,٦٠ وقيم معامل ثبات درجات الاختبارات تتراوح بين ٠,٨٠ ، ٠,٨٤ .

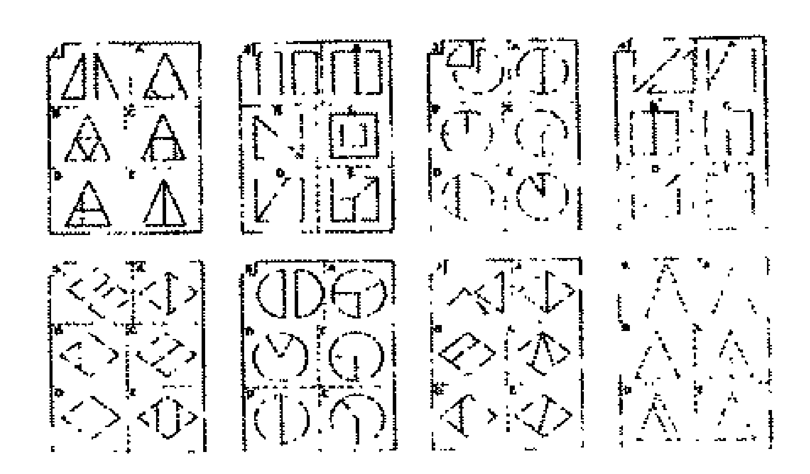
(٢) اختبار مينيسوتا للعلاقات المكانية

Minnesota Spatial Relations Test

أعد هذا الاختبار ترابيو Trabue وزملاؤه. وصدرت طبعته الثانية المُعدَّلة عام ١٩٦٩التى أعدها ليكرت Likert ، وكواشا Quasha. لقياس المكونة الرئيسة الثانية من مكونات الاستعداد الميكانيكى. فإدراك العلاقات المكانية هو القدرة على الحكم والتعامل عقليا مع الأشياء فى الفراغ، ومعرفة تأثير إعادة تنظيمها ، والتوصل إلى العلاقات القائمة بينها فى وضعها الجديد، أى أنه يتعلق بالتوجيه المكانى Spatial Orientation أو التصور البصرى المكانى. والاختبار بعد تعديله يُطلق عليه اختبار لوحات الورق Revised Minnesota Paper Form Board Test (MPFB)

ويشتمل على ٦٤ رسمًا هندسيًا لأجزاء غير مرتبة وُضعت كل مجموعة منها فى ستة خلايا من لوحات مستطيلة، وبعض هذه الأجزاء مقلوبة أو تم تدويرها، واللوحات متدرجة الصعوبة.

ويطلب من الفرد اختيار الشكل الهندسى الصحيح بعد تجميعه من بين بدائل أشكال معطاة كما هو موضح بشكل (١٠ ـ ٣) التالى (Guion , 1965):



شكل (١٠ ـ ٣) يوضح بعض مفردات اختبار العلاقات المكانية

ومن الجدير بالذكر أن معظم بطاريات الاستعدادات الخاصة وكثيراً من اختبارات الذكاء تشتمل على اختبار فرعى يقيس العلاقات المكانية، نظراً لأن الدراسات أوضحت

أن هذه القدرة منبئ جيد بالنجاح فى المهن والمجالات الدراسية المختلفة. غير أن هذا الاختبار يُعد من الاختبارات الجيدة لقياس هذه العلاقات، حيث أشارت إليه معظم الدراسات التى تناولت الاستعداد الميكانيكى. فهو يفيد فى التنبؤ بالنجاح فى المجالات الهندسية والتصميم المعمارى وتصميم الآلات وغيرها.

والاختبار يناسب طلاب الصفوف الأخيرة من التعليم الثانوى وكذلك الراشدين، وأُعدَّت معايير مئينية خاصة بهؤلاء الطلاب، ومعايير أخرى للمتقدمين لهذه المهن والمتدربين والمهندسين.

وقد أُعدَّت صيغتان متكافئتان للاختبار، وتتراوح قيم ثبات درجات الاختبار بين ٠,٨٠ ، ٠,٨٥ إذا استُخدمت أى من الصيغتين، بينما تراوحت القيم بين ٠,٨٩ ، ٠,٩٢ إذا استُخدمت الصيغتان معًا ؛ لأن ذلك يؤدى إلى إطالة الاختبار. وأُجريت دراسات متعددة للتحقق من الصدق التنبؤى للاختبار فى مهن مختلفة تتعلق بالمجالات الميكانيكية وغير الميكانيكية التى تتطلب إدراك العلاقات المكانية. وقد تراوحت قيم معامل الصدق التنبؤى للمجالات الميكانيكية بين ٠,٣٨ ، ٠,٥٩ ، وللمجالات غير الميكانيكية مثل الفنون بين ٠,٢٨ ، ٠,٥٧ ، كما وُجد أن ارتباط درجات هذا الاختبار بتقديرات الطلاب فى كليات الهندسة بلغ ٠,٣٥ ، وارتباطه بدرجات اختبار الفهم الميكانيكى تراوح بين ٠,٢٢ ، ٠,٥١ ، وباختبار تصميم المكعبات الذى يُعد أحد اختبارات مقياس ويكسلر - بلفيو ٠,٤٠ .

اختبارات الاستعدادات النفسية الحركية :

يلاحظ أن هناك بعض اختبارات الورقة والقلم التى تقيس المعلومات والخبرات الميكانيكية، مثل اختبار أورورك للمعلومات الميكانيكية O'Rourke Mechanical Information Test الذى يتطلب فى جزئه الأول أن يزاوج الفرد بين صور الأدوات اليدوية وصور الأشياء التى تستخدم معها. وفى الجزء الثانى يشتمل على مفردات اختيار من متعدد تقيس معلومات تتعلق بالأدوات واستخداماتها فى مجالات النجارة والكهرباء والطلاء والطباعة وغيرها.

كما أن هناك اختبارات الاستعدادات النفسية الحركية التى تقيس المهارات اليدوية المختلفة Manual Dexterity ، مثل اختبارات حركات الأطراف والأصابع والعضلات فى أداء مهام تجميعية معينة بمهارة وسرعة ودقة. ولكن ينبغى التمييز بين الاستعداد الميكانيكى والاستعداد اليدوى، فعلى الرغم من أن كليهما يتطلب استخدام الأدوات والآلات والمعدات الميكانيكية، إلا أن الاستعداد الميكانيكى-كما سبق أن ذكرنا-يعتمد

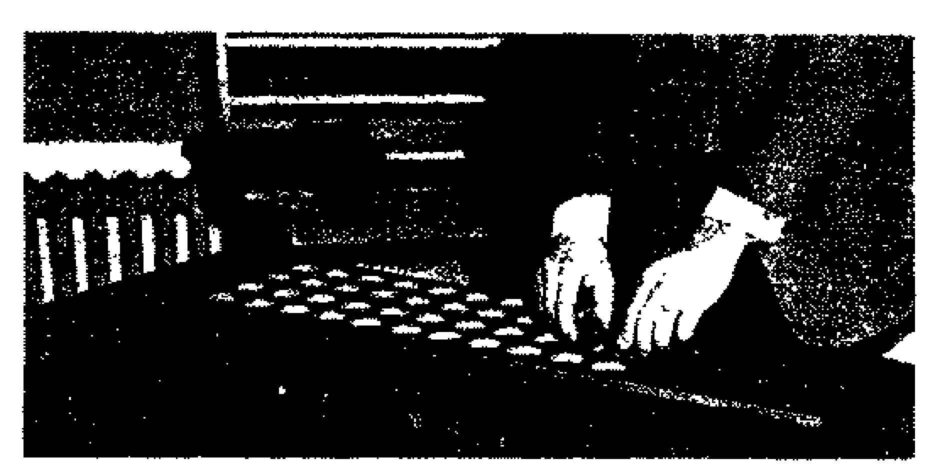
على عمليات عقلية رمزية تؤدى إلى اتخاذ قرارات تتعلق بما يجب أن تفعله الأيدى والأدوات والآلات، وربما يجرى الفرد هذه العمليات ببطء ومع هذا ينجح فى مهنته الميكانيكية.

فما يهمنا هنا هو قدرة الفرد على تشخيص الأعطال، وكيفية إصلاحها. أما الاستعداد اليدوى فهو متطلب أساسي فى خطوط التجميع الآلية التى لا تحتاج إلى اتخاذ قرارات معينة، حيث إن الفرد يعطى تعليمات محددة وواضحة بما يجب أن يفعله، وعليه أن يؤدى المهام بمهارة وسرعة ودقة، وهذا يعتمد على مهارته اليدوية وليس على فهمه الميكانيكى. وهناك اختبارات متعددة تقيس مدى سرعة ودقة حركات الذراع واليد والأصابع فى التعامل مع الأدوات والأشياء والأجزاء الدقيقة، وتختلف هذه الاختبارات باختلاف درجة دقة الأجزاء المراد التعامل معها.

(١) اختبار مينيسوتا لمعدل الحركات

Minnesota Rate of manipulation Test (١٩٦٩):

وهو اختبار شائع الاستخدام يتطلب حركات اليد والذراع، وهو عبارة عن لوحة بها ٦٠ ثقبًا فى أربعة صفوف، ويوجد ٦٠ قرصًا صغيرًا اسطوانيًا متماثلاً إحدى قاعدتى كل منها ملون باللون الأحمر، والقاعدة الأخرى باللون الأخضر، كما هو مبين بشكل (١٠ ـ ٤) التالى (Guion , 1965):



شكل (١٠ ـ ٤) يوضح بعض مفردات اختبار معدّل الحركات

والمهام التي يتطلبها الاختبار هى . تدوير الأقراص أو تعديل وضعها وإدخالها فى الثقوب، أو إدخال الأقراص فى الثقوب من وضع جانبى بالنسبة للوحة، وإخراج الأقراص من الثقوب، واستخدام يد واحدة فى تعديل وضع الأقراص وإدخالها فى الثقوب، ثم استخدام اليدين معًا فى إجراء ذلك.

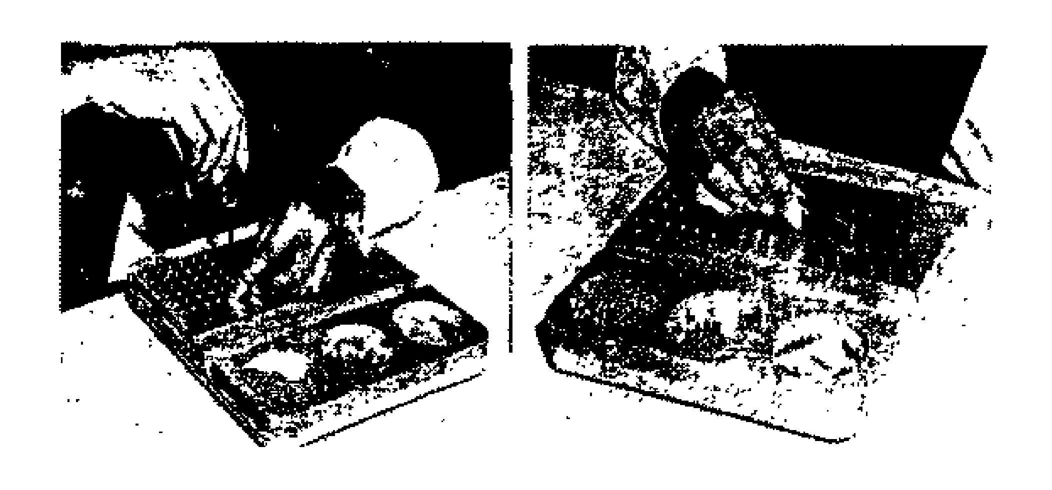
ومقدار الزمن الذى يستغرقه الفرد فى إدخال جميع الأقراص فى الثقوب يُمثل درجته في الاختبار، وقد استمدت معايير الاختبار من تطبيقه على عينة اشتملت على ١١٠٠٠ من العاملين وغير العاملين من الشباب، وعينة أخرى من غير العاملين من الكبار اشتملت على ٣٠٠٠ فرد، ويقارن الزمن الذى استغرقه الفرد بهذه المعايير. وقدم دليل الاختبار بعض المؤشرات عن الصدق التنبؤى للاختبار فى مواقف تطبيقية متعددة.

(٢) اختبار كراوفورد لمهارة التعامل مع الأجزاء الدقيقة

:Crawford Small Parts Dexterity Test

يقيس هذا الاختبار المهارة الدقيقة وتآزر العين واليد، ويشتمل الاختبار على جزئين: أحدهما يتطلب من الفرد استخدام ملقط فى إدخال دبابيس فى ثقوب رفيعة ووضع غطاء فوقها، والآخر يتطلب وضع مسامير لولبية دقيقة فى ثقوب بها خيوط وربطها بمفك.

ويوضح شكل (١٠ ـ ٥) التالى صورة الجهاز المستخدم فى ذلك :(Guion , 1969)



شكل (١٠ ـ ٥) يوضح بعض مفردات اختبار مهارة التعامل مع الأجزاء الدقيقة

وتعتمد معايير الاختبار على الزمن الذى يستغرقه الفرد لإنجاز المطلبين، واستمدت من عينات اشتملت على عشر مجموعات من العاملين والمتقدمين للالتحاق بالعمل. وقد تبين أن قيمة معامل الارتباط بين درجات جزئى الاختبار ٠,٤٠ تقريباً، فى حين أن ثبات درجات الاختبار بطريقة التجربة النصفية بلغت قيمته ٠,٨٥ تقريباً.

والصدق التنبؤى للاختبار فى المجالات المهنية التى تتطلب مهارة يدوية دقيقة يحتاج إلى مزيد من الدراسات نظرًا لأن ما أشار إليه دليل الاختبار فى هذا الشأن لا يُمكننا من تقييم صدق الاختبار.

الاستعداد الكتابى :

يتضمن الاستعداد الكتابى Clerical Aptitude المهارات المكتبية المتعددة المتعلقة بأعمال الكتابة على الآلة الكاتبة أو الحاسوب، وحفظ السجلات وتحديثها، والاختزال وتدقيق الحسابات وغير ذلك.

والاستعداد الكتابى ليس قدرة واحدة وإنما يتضمن قدرات متعددة. وعلى الرغم من أن هناك أفرادًا يؤدون أحد هذه الأعمال وليس جميعها، إلا أن هناك اختبارات تقيس الاستعداد الكتابى العام، واختبارات أخرى تقيس قدرات خاصة نوعية. فالاستعداد الكتابى العام ينطوى على قدرات تماثل القدرات التى ينطوى عليها الذكاء العام.

لذلك تشتمل الاختبارات التى تقيس هذا الاستعداد على مفردات مماثلة للمفردات التى تشتمل عليها اختبارات الذكاء العام، مثل القدرات اللفظية والعددية، غير أنها تشتمل أيضًا على مفردات تقيس المهارة اليدوية، وسرعة إدراك أوجه الشبه والاختلاف بين الكلمات والأعداد. كما أن بطارية اختبارات الاستعدادات الفارقة (DAT) التى أشرنا إليها فيما سبق تتضمن بعض الاختبارات الفرعية المتعلقة بالاستعداد الكتابى.

وسوف نُقدم فيما يلى أمثلة لبعض اختبارات الاستعداد الكتابى العام، والاستعدادات النوعية شائعة الاستخدام المتعلقة به:

اختبارات تقيس الاستعداد الكتابى العام :

(١) اختبار الاستعداد الكتابى العام (١٩٧٢)

General Clerical test (GCT):

يُعد هذا الاختبار بطارية من الاختبارات تشتمل على تسعة أجزاء يحصل الفرد فيها على ثلاث درجات فرعية؛ إحداها تتعلق بسرعة ودقة أداء المهام الكتابية الروتينية، وتتطلب مراجعة الأعمال الورقية، وترتيب الحروف، والثانية تتعلق بالمهام الحسابية التى تُعد ضرورية لبعض الأعمال المكتبية، وتتطلب إجراء عمليات حسابية، وتحديد أخطاء الحسابات، والاستدلال الحسابى، والثالثة تتعلق بالجانب اللفظى أى المهارات اللغوية المهمة لأداء المهام الكتابية، وتتطلب مهارات الهجاء، ومهارات فهم المقروء، والكلمات، والنحو. ويمكن الحصول على درجة الاستعداد الكتابى العام بجمع الدرجات الثلاث.

وقد أُعدَّت معايير مئينية مستقلة لكل من الأجزاء الثلاثة لكل من الذكور والإناث استنادًا إلى جماعات مرجعية متعددة تشمل المتقدمين للالتحاق بوظائف كتابية، وطالبات نهاية مرحلة التعليم الثانوى التجارى، والعاملين فى الوظائف الكتابية المختلفة.

وتتميز البطارية بسهولة تطبيقها وتصحيحها ، ويستغرق تطبيقها ٤٧ دقيقة تقريبًا، وحدِّد زمن مستقل لكل جزء من الأجزاء الثلاثة، وتراوحت قيم معامل ثبات درجات كل من هذه الأجزاء بين ٠,٨٢ ، ٠,٩١ ، وقيمة معامل ثبات الدرجات الكلية ٠,٩٤ واستُخدمت طرق مختلفة للتحقق من صدق البطارية فى التنبؤ بالنجاح فى الأعمال الكتابية بعامة وكذلك الصدق التلازمى، وكانت النتائج إيجابية.

كما وجد أن درجات البطارية ترتبط بدرجات أحد اختبارات الذكاء العام ارتباطًا موجبًا تراوحت قيمته بين ٠,٥٤ ، ٠,٨٣ ، مما يشير إلى العلاقة بين الاختبارين . غير أن هذه العلاقة تكون منخفضة فى حالة اختبارات الاستعدادات الكتابية النوعية التى سنوضحها بعد قليل.

(٢) اختبار مينيسوتا للاستعداد الكتابى

Minnesota Clerical Test (MCT):

يُعد هذا الاختبار من أكثر اختبارات الاستعدادات الخاصة استخدامًا حيث اعتنى ببنائه وتقنينه ومراجعته وأُجريت حوله العديد من الدراسات، وقد أعده أندرو Andrew، وباترسون Paterson. ويمكن الإفادة من هذا الاختبار فى التنبؤ بالنجاح فى مهن كتابية مختلفة، مثل العمل المكتبى العادى، وأعمال الفحص والمراقبة الإدارية. فهذه الأعمال تتطلب سرعة الإدراك البصرى، والسرعة والدقة فى التعامل مع الأعداد والحروف وغيرها من الرموز.

لذلك يهدف الاختبار لقياس سرعة القراءة، والفهم، واستخلاص معلومات من التقارير والمراسلات والأسماء والأعداد. ويتكون الاختبار من جزئين: أحدهما يتعلق بمراجعة الأعداد والآخر يتعلق بمراجعة الأسماء ويشتمل كل من الجزئين على أزواج من الأعداد أو الأسماء (قمنا بتعريب الأسماء) كما يلى :

المقارنة بين الأعداد		المقارنة بين الأسماء	
٣٤٥١٢	٣٤٥١٢	طارق شاهين	طارق شاهين
٤٧٣٠٦	٤٧٣٠٦	حسن السيد	حسين السيد
٨٩٧٠٥	٨٩٥٠٧	مدينة نصر ، القاهرة	مدينة نصر القاهرة
١٣٧٤٢٨	١٧٣٤٢٨	الشركة التجارية ش م م	الشركة التجارية ش.م.م

ويُطلب من الفـرد فحص هذه الأزواج ، ووضع علامة على الخـط الفاصل بين مفردتى كل زوج منهـا إذا كان هناك تطابق بين المفردتين ولا تـوضع أى علامة إذا كانا مختلفـتين. وتتكون القائمة من ٢٠٠ زوج من الأعـداد، ٢٠٠ زوج من الأسماء على الترتيب، وتُقدر درجة لكل جزء، ولا تُجمع درجات الجزئين.

ويتميز الاختـبار بسهولة التطبيق والتصحـيح، ويستغرق تطبيقه ١٥ دقيـقة تقريبًا حيث إنـه يعتمـد على السرعـة، والاختبار ينـاسب طلاب المرحلـة الثانوية. وتتعدد المعايير المئينية للاختبار ، إذ توجد مـعايير منفصلة للذكور والإناث، وللفرق الدراسية بالمرحلة الثانوية، وللمناطق الجغرافية، وللمـتقدمين للأعمال الكتابية ، ولبعض المهن الكتابية المـتنوعة. وقيم معاملات ثبات درجـات الاختبار بطريقة إعادة تطبـيق الاختبار تراوحت بين ٠,٧٧ ، ٠,٩١ . غير أن ارتـباط درجات الاختبار بتقديـرات المعلمين والمشرفين في مجالات العمل الكتابية تراوحت قيمه بين ٠,٢٨ . ٠,٦٥ . والارتباط بين درجات الجزء المتعلق بمراجعة الأعداد ودرجات اختبار الاستعداد الأكاديمى بلغت قيمته ٠,١٠ ، بينما بلغت قيمة الارتباط بين درجات الجزء المتعلق بمراجعة الأسماء، ودرجات اختبار الاستعداد الأكاديمـى ٠,٣٥ ، وكذلك وُجد أن قيمة الارتباط بين هذا الاختبار واختبار الاستعداد الكتابى العام الذى أشرنا إليه فيما سبق ٠,٣٦ .

ويتضـح من ذلك أن السرعة الإدراكـية تُعد أحـد العناصر الأسـاسية لكثـير من الأعمال الكتابية، ولكن ينبغى أيضًا قيـاس عناصرأخرى مثل الفهم اللفظى، والعمليات الحسابـية. وقد أُعدت اختبـارات أخرى لقياس هذه العنـاصر، مثل الاختبار الـمـسحى لاستعداد العاملين Employee Aptitude Survey الذى يقيس الفهم اللـفظى المتمثل فى القدرة على استخدام الكلمات فى التواصل والتـفكير والتخطيط. ويتضمن الاختبار أيضًا مفردات لـقياس القدرة العددية المُتـمثلة فى التعامل بسهـولة مع الأعداد، وإجراء العـمليات الـحسابـية البـسيطة بـسرعـة ودقة . وكذلـك اختبـار فلاناجـان لتصـنيف الاستعدادات Flanagan Aptitude Classification Test الذى يشتمـل على مفردات لقياس مهارات التواصل، وبخاصة فيما يتعلق بتركيب الجمل ، والنحو.

اختبارات تقيس استعدادات كتابية نوعية :

اختبارات الاستعداد المتعلق بالاختزال والآلة الكاتبة:

يوجد العديد من الاستـعدادات التى تتعلق بأعمال أو مهن خـاصة، مما يتطلب بناء اختبارات لقياس هذه الاستعدادات استنادًا إلى إجراءات تحليل العمل ومتطلباته التى سبق أن أوضحناها بـالتفصيل. فالأعمال الكتـابية المتعلقة بالكتابـة على الآلة الكاتبة،

والاختزال Stenographic Jobs تتطلب قدرات خاصة ومهارات متعددة. فتعلم الاختزال يعتمد على اكتساب نوع جديد من التمثيل الرمزى اللغوى، وبذلك يتضمن المهارات اللفظية. وتعلم الآلة الكاتبة يعتمد على حدة البصر، ومهارة الأصابع، وتآزر العين والأصابع، وغير ذلك من المهارات. وتلعب السرعة دوراً مهمًا فى تقييم الأداء فى هذه الأعمال.

ومن بين الاختبارات التى تقيس هذه الاستعدادات اختبار الكفاءة فى أعمال الاختزال Seashore - Bennett Stenographic Profeciency Test (١٩٥٦).

وتُعرض المعلومات على الفرد المختبر عن طريق الحاسوب أو شرائط التسجيل، وهذه المعلومات عبارة عن خطابات تجارية غير فنية، وذلك لتقييم كفاءته فى أعمال الاختزال. ويتكون الاختبار من صيغتين متكافئتين، وتتضمن كل منهما خمسة خطابات. وتتباين فى عدد فقراتها وسرعة عرضها. وبعد إملاء الخطابات فى مدة ١٥ دقيقة يقوم الفرد بنسخها على الآلة الكاتبة فى مدة ٣٠ دقيقة. ويُستخدم هذا الاختبار فى انتقاء العاملين فى مجال الاختزال وأعمال السكرتارية.

اختبارات الاستعداد المتعلق ببرمجة الحاسوب :

تتطلب الأعمال المتعلقة ببرمجة الحاسوب قدرات خاصة أو مهارات معينة ينبغى توافرها فى المتقدمين لهذه الأعمال، غير أن الاختبارات التى تقيس هذه القدرات محدودة، ومن أمثلتها بطارية اختبارات استعداد برمجة الحاسوب Computer Programming Aptitude Battery الذى أعدها بالورمو Palormo .

وتشتمل البطارية على خمسة اختبارات فرعية تتعلق بفهم الكلمات، وتقريب الأعداد، وسلاسل الحروف، والاستدلال، وتصميم الأشكال التخطيطية. ويستغرق تطبيق البطارية ٧٥ دقيقة، وتُقارن درجة الفرد فى كل من الاختبارات الخمسة ودرجته الكلية بمعايير مئينية.

ونظراً لانتشار استخدام الحاسوب فى الأعمال الكتابية فإن الاهتمام أصبح مُوجهًا لبناء اختبارات استعداد فى هذا المجال بديلاً عن الاختبارات المتعلقة بالآلة الكاتبة.

الاستعداد الموسيقى :

يُعد الاستعداد الموسيقى سمة مركبة من مجموعتين من المكونات تتعلق إحداهما بالذوق والإلهام والانفعال والإبداع، وتتعلق الأخرى بالجوانب الآلية والحسية والإدراكية. وكل من المجموعتين ربما يكون ضرورياً للأداء الموسيقى الجيد ولكنهما

غير كافيين للنجاح فى المهن الموسيقية. ومع هذا فإن اختبارات الاستعداد الموسيقى المتوافرة تركز على قياس المجموعة الثانية من المكونات، وذلك لأنه يمكن تحديدها تحديداً دقيقًا وقياسها بسهولة.

غير أن كثيراً من هذه الاختبارات أُعد منذ مدة طويلة، والاهتمام ببناء اختبارات جديدة لقياس الاستعداد الموسيقى يبدو محدوداً. وتعتمد الإجابة عن مفردات معظم هذه الاختبارات على الحكم الشخصى للفرد المُختبر، حيث لا يُطلب منه العزف الفعلى على آلة موسيقية معينة، ولكنه يستمع إلى لحن معين ويحكم على هذا اللحن من حيث النغمة والإيقاع وحدَّة الصوت وغير ذلك من الخصائص، فالعزف على الآلات الموسيقية يتطلب إلى جانب ذلك مهارات نفسية حركية.

وعلى الرغم من أن كثيراً من اختبارات الاستعداد الموسيقى تقيس المكونات الأساسية للاستعداد الموسيقى الناتجة عن تحليل هذا الاستعداد تحليلاً سيكولوجيًا، إلا أن بعضها يقيس المحتوى الموسيقى فى تكامله وشموله. وسوف نقدم فيما يلى أمثلة لبعض هذه الاختبارات :

(١) مقاييس سيشور للاستعداد الموسيقى

:Seashore Measures of Musical Talents

وتُعد هذه المقاييس أقدم مقاييس الاستعداد الموسيقى التى استندت إلى البحوث الرائدة التى أجراها سيشور وزملاؤه بجامعة أيوا الأمريكية (Seashore , 1938)، وأعيد مراجعتها عامى ١٩٥٧ ، ١٩٦٠ ، حيث أُجريت تعديلات فى التعليمات، كما طوِّرت أجهزة التسجيل لكى تتماشى مع المستحدثات التكنولوجية.

واعتمد بناء المقاييس على عدد من العناصر المتعلقة بالموهبة الموسيقية، حيث يُقدم للفرد المختبر سلسلة من المهام ويطلب منه تمييز الفروق بين أزواج المثيرات، وتُقدم هذه المثيرات فى ستة اختبارات منفصلة مسجلة على شريط ليزر. وتقيس هذه الاختبارات درجة الصوت، وعلوه، والإيقاع، والزمن، وسرعة النغمة، وذاكرة النغم.

ويمكن تطبيق هذه المقاييس على الأطفال من الصف الرابع الابتدائى إلى الرشد. ويستغرق تطبيق المقاييس ساعة واحدة تقريبًا بمعدل عشر دقائق لكل اختبار. وقد أكد سيشور ضرورة تصحيح كل اختبار على حدة، ولا تُجمع درجات الاختبارات، وأعدت معايير مئينية لكل اختبار للفرق الدراسية المختلفة، وتراوحت قيم معامل ثبات

درجات الاختبارات الستة باستخدام طريقة التجزئة النصفية بين ٠,٥٥ ، ٠,٨٥ . وأوضحت نتائج بعض الدراسات أن ارتباط درجات الاختبارات بمحك التدريب الموسيقى تراوحت قيمته بين ٠,٣٠ ، ٠,٤٠ ، غير أن بيانات الصدق التنبؤى للمقاييس غير كافية.

(٢) بطارية وينج المقننة للذكاء الموسيقى

:Wing Standaradized Tests of Musical Intelligence

أعد وينج Wing هذه البطارية فى إنجلترا عام ١٩٤١ وأعيد مراجعتها وصدرت عام ١٩٦٢ بلغات ثلاث:الإنجليزية والفرنسية والألمانية . وتتميز هذه البطارية عن مقاييس سيشور فى أنها لم تعتمد على قياس مكونات الاستعداد الموسيقى، ولكن اعتمدت على المحتوى الموسيقى ككل.

ويمكن تطبيق البطارية ابتداء من عمر ٨ سنوات، ويُستخدم فيها موسيقى البيانو على شريط تسجيل، وتتكون البطارية من سبعة اختبارات فرعية هى : تحليل الأوتار من حيث عدد النغمات التى تُصدرها، وتمييز النغمات، وذاكرة النغم، وائتلاف النغم، ووحدة الصوت، والإيقاع، والتعبير. وتختلف هذه البطارية أيضًا عن مقاييس سيشور فى أن الاختبارات الأربعة الأخيرة تتطلب من الفرد تقييم الجوانب الذوقية للنغمات المختلفة. ويستغرق تطبيق البطارية ساعة واحدة تقريبًا، ويمكن تطبيقها فى جلستين.

وأعدت معايير منفصلة للأعمار من ٨ إلى ١٧ عامًا للدرجات الكلية ودرجات الاختبارات الفرعية الثلاثة الأولى التى اعتُبرت بمثابة بطارية قصيرة تناسب الأطفال الصغار وكأداة فحص أولية للاستعداد الموسيقى. وهذه المعايير عبارة عن تقديرات بالحروف وكذلك الأعمار الموسيقية. وتراوحت قيم معامل ثبات درجات الاختبارات الفرعية بين ٠,٧٠ للأطفال الصغار، و ٠,٩٠ للأطفال الأكبر عمرًا. كما أن صدق البطارية فى التنبؤ بمحك الأداء الموسيقى يحتاج إلى مزيد من الدراسات، على الرغم من أنه وُجد ارتباط بين درجات البطارية وتقديرات معلمى الموسيقى لطلابهم بلغت قيمته ٠,٦٠ أو أعلى. وقد نالت هذه البطارية اهتمام كثير من معلمى الموسيقى لأنهم يرون أن محتواها يتضمنُ المهارات ذات الأهمية فى التدريبات الموسيقية.

لذلك يمكن استخدامها فى التوجيه وانتقاء الطلاب الذين يرغبون فى متابعة التدريبات الموسيقية المتقدمة. ومن الجدير بالذكر أن هناك بطارية أخرى تعرف باسم «بروفيل الاستعداد الموسيقى Musical Aptitude Profile» أعدها جوردون عام ١٩٦٥ وأجرى مراجعتها عام ١٩٦٧ ، وتعتمد على آلات موسيقية مختلفة، وتشتمل على سبعة

اختبارات صُنفت في ثلاث مجموعات هي : تصور النغمات ، وتصور الإيقاع، والحس الموسيقى.

ومن هذا يتضح أن الصدق التنبؤى لبطاريات الاستعداد الموسيقى يتطلب مزيداً من البحث، وربما يُفضل استخدام اختبارات أخرى إلى جانب هذه البطاريات للحصول علي تقدير أفضل لهذا الاستعداد.

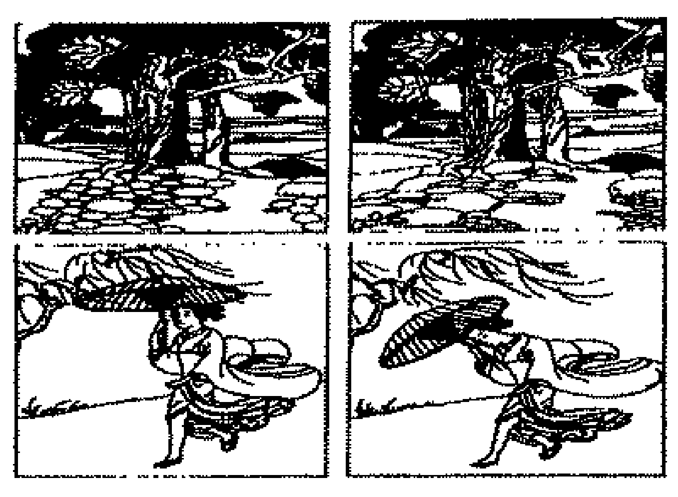
الاستعدادات الفنية :

إن قياس الاستعدادات الفنية Art Aptitudes يعد من المجالات التى لم تحظ بنفس القدر من الاهتمام الذى حظى به قياس الاستعدادات الأخرى التى أوضحناها فيما سبق.

ولعل هذا يرجع إلى صعوبة الفصل بين الاستعدادات الفنية والتحصيل ، حيث إن بعض اختبارات هذه الاستعدادات تقيس المعلومات والخبرات المتعلمة فى المجالات الفنية. كما أن الاهتمام بانتقاء وتصنيف وتسكين الأفراد فى المدارس والمصانع والمجالات العسكرية نال معظم اهتمام علماء القياس مما أدى إلى قلة الاختبارات فى مجال الفنون، وساعد على ذلك أيضًا تعدد وتباين القدرات التى يتطلبها هذا المجال. كما أن الفرد الذى يستطيع تذوق الفن أو نقده ليس بالضرورة لديه الاستعداد الذى يُمكنه من الإنتاج الفنى.

لذلك ربما يكون من الضرورى التمييز بين مقاييس الاستعدادات المتعلقة بالحكم الذوقى Aesthetic Judgment، والإنتاج الفنى Aesthetic Production التى تعتمد على عينات فعلية من الأعمال الفنية.

ويوضح شكل (١٠ ـ ٦) التالى بعض مفردات هذه الاختبارات (Nunnally , 1975):



شكل (١٠ ـ ٦) يوضح بعض مفردات اختبارات التقييم الفنى

وسوف نقدم فيما يلى أمثلة لبعض مقاييس كل من النوعين.

أولاً: اختبارات الاستعداد المتعلق بالحكم الذوقى:

(١) اختبارات ماير للتقييم الفنى Meier Art Judgment Tests:

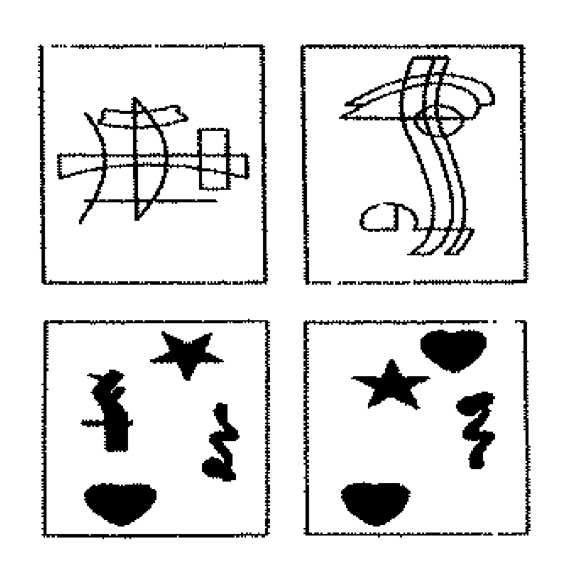
تُعد من الاختبارات المعروفة فى هذا المجال، وقد أعدها ماير Meier فى طبعتها الأولى عام ١٩٢٩، وأعيد طباعتها عام ١٩٦٣، وتشتمل الاختبارات على ١٠٠ زوج من اللوحات الفنية الشهيرة غير الملونة، وإحدى لوحات كل زوج تُعد لوحة أصلية واللوحة الأخرى بها بعض التعديلات البسيطة. ويُطلب من الفرد تحديد أوجه الاختلاف بينهما، ويقرر أيهما يفضل. فالاستعداد المتعلق بالحكم الذوقى ينبغى أن يتوافر لدى الناقد الفنى، أما الاستعداد المتعلق بالإنتاج الفنى فينبغى أن يتوافر لدى الفنان نفسه. وهذا الإنتاج ربما يتطلب أيضًا الحكم الذوقى، ولكن العكس لا يكون صحيحًا.

وقد أُعدت معايير مئينية لطلاب المرحلة الإعدادية والثانوية، وتراوحت قيم معامل ثبات درجات الاختبارات بطريقة التجزئة النصفية بين ٠,٧٠ ، ٠,٨٤ ، وتراوحت قيم معامل الارتباط بينها وبين تقديرات معلمى المجالات الفنية بين ٠,٤٠ ، ٠,٧٠ ، وهى قيم مقبولة. ومن الجدير بالذكر أن ماير Meier أعد اختبارًا للإدراك الجمالى Aesthetic Perception عام ١٩٦٣ لقياس الفروق الفردية فى إدراك القيمة الجمالية للطرق المختلفة فى تكوين لوحة فنية ، ويشتمل الاختبار على ٥٠ لوحة فنية لكل منها أربعة بدائل ، ويُطلب من الفرد ترتيب اللوحات وفقًا لقيمتها الجمالية كما يُدركها.

(٢) اختبارات جريفز للحكم على التصميمات الفنية

Graves Design Judgment Tests ،

أعد جريفز Graves هذا الاختبار عام ١٩٤٨وأعيد طباعته عام ١٩٥١ . ويختلف عن اختبار ماير Meier فى أنه لا يُقدم للفرد بدائل أعمال فنية شهيرة ولكنه يعتمد على التصميمات المجـردة ، حيث يشتمل على ٩٠ مجموعة من التصمـيمات المجردة مما يجعل الاخـتبار لا يعتمـد على القيم الفنـية المتعارف عـليها. وتتكون كل مـفردة من نسختـين أو ثلاثة من التصميـم نفسه ولكنها تُخـالف مبادئ الفن الذوقي المـقبول من حيث الوحدة والتوازن والتماثل أو غير ذلك من مظاهر الذوق الفنى ، ويطلب من الفرد أن يقرر أيهـا أفضل . والاختبار غيـر موقوت ، ولكن تطبيـقه يستغرق نصـف الساعة تقريبًا. ويوضح شكل (١٠-٧) التالى بعض مفردات هذا الاختبار (Nunnally , 1965):



شكل (١٠ - ٧) يوضح بعض مفردات اختبار الحكم على التصميمات الفنية

وقد تراوحت قـيم معامل ثبـات درجات الاختبـار بطريقة التـجزئة النصفـية بين ٠,٨٠ ، ٠,٩٣ ، ويتطلب الاختبار مزيداً من دراسات الصدق التنبؤى.

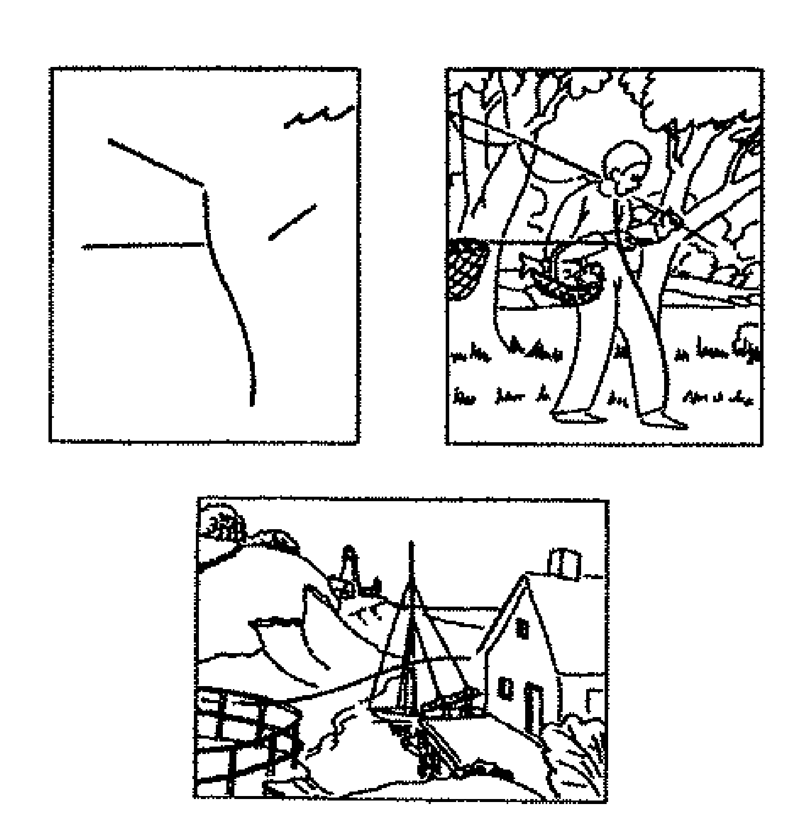
ثانيا : اختبارات الإستعداد المتعلق بالإنتاج الفنى :

تعتمد معظم اختبارات الاستعداد المتعلق بالإنتاج الفنى على عينات من الأعمال الفنية مما يتطلب العناية بانتقاء هذه العينات، وتتأثر درجاتها بالتدريب المنتظم، ويوجد عدد محدود من هذه الاختبارات لعل أهمها:

قائمة هورن للاستعداد الفنى Horn Art Aptitude Inventory:

أعد هورن Horn هذه القائمة التى تشتمل على عينات من الأعمال الفنية ، وتتكون من جزئين: أحدهما يتطلب من الفرد عمل رسم تصويرى لعشرين شيئًا مألوفًا مختلفًا مثل شجرة أو كتاب أو سكين، على أن يتم العمل فى مدة قصيرة لا تزيد عن عشر ثوان ، والآخر يتطلب التصور، ويشتمل على ١٢ مستطيلاً يتضمن كل منها عدة خطوط يسترشد بها الفرد فى رسم صورة ممثلة.

ويوضح شكل (١٠ ـ ٨) التالى أمثلة لبعض مفردات الجزء المتعلق بالتصور فى هذه القائمة (Anastasi , 1976):



شكل (١٠ ـ ٨) يوضح بعض مفردات قائمة الاستعداد الفنى

وقد أشار دليل القائمة إلى أمثلة للرسم الممتاز والمتوسط والضعيف، وقدم بعض محكات تقدير الدرجات. وقد وُجد أن معامل ارتباط درجات القائمة ومتوسط

تقديرات معلمي المجالات الفنية لعينة صغيرة من طلاب المرحلة الثانوية ٠,٥٣ ، غير أن ثبات درجات القائمة وصدقها يحتاج إلى مزيد من الدراسات.

الابتكارية :

شهدت العقود التي تلت الحرب العالمية الثانية تطورات ملحوظة في كثير من المجالات المهنية العليا مثل المجالات التقنية، ومجالات البحث العلمي، والمجالات القيادية التنفيذية، ومجالات الفنون مما أدى إلى تنشيط بحوث القياس النفسي والتربوي لمواجهة المشكلات الجديدة الملحة التي نجمت عن هذه التطورات.

فعلى الرغم من أن اختبارات الذكاء والاستعدادات الخاصة التي تناولناها فيما سبق يمكن باستخدامها التنبؤ إلى حد ما بأداء الأفراد في هذه المجالات المهمة، إلا أن نطاق هذه الاختبارات لم يمتد بحيث يشمل القدرات الأكثر أهمية المتعلقة بالتفكير الابتكاري Creative Thinking وما يتضمنه من أنشطة عقلية غير مألوفة. فاختبارات الذكاء التقليدية تتطلب من الفرد تطبيق ما تعلمه في موقف جديد، أو استنباط قاعدة معينة من أمثلة معطاة لفظية أو عددية أو شكلية، ويُعيد تطبيق هذه القاعدة. وعادة تكون هناك إجابة واحدة لكل مفردة من مفردات الاختبار، وتُحسب الدرجة الكلية في الاختبار بإيجاد مجموع درجات الإجابات الصحيحة.

لذلك فإن إجابة الفرد عن هذه الاختبارات تتطلب تفكيراً تقاربياً محدداً Convergent Thinking. فمثلاً إذا طُلب من الفرد أن يُحدد العدد التالي في السلسلة (٢، ٤، ٨، ١٦) فإن عليه أن يدرك أن كل عدد منها ضعف الذي يسبقه، وبذلك يستنتج أن العدد التالي (٣٢) ، غير أن الفرد ربما يتبع نمط تفكير مختلف، إذ ربما يلاحظ أن خانة العشرات في الأرقام الثلاثة الأولى (صفر) ، حيث يمكن كتابتها (٠٢، ٠٤، ٠٨) على الترتيب. وبعد إجراء هذا التحويل على المفردة الاختبارية، فإن الأعداد الثلاثة التالية يمكن أن تتبع قاعدة مماثلة لذلك. ونظراً لأن خانة العشرات واحد صحيح في العدد التالي (١٦)، فإن العدد الذي يكمل السلسلة يمكن أن يكون (١٢)، وذلك لأن الرقم الأول (٢) هو رقم الآحاد في العدد (٣٢). وهذا النمط من التفكير على الرغم من صحته واختلافه عن نمط التفكير السابق، إلا أن الفرد يحصل على الدرجة صفر في هذه المفردة، ويُعد هذا النمط تفكيراً تباعدياً أو منطلقا Divergent Thinking. أي أنه في إطار اختبارات الذكاء التقليدية تُعد هذه الإجابة خطأ لأنها إجابة غير مألوفة، وذلك لأن مفردات هذه الاختبارات تُعد وحدات متكاملة في ذاتها،

ولا يجب على الفـرد الذى يجيب عنهـا إجراء أى تعديلات أو تحويـرات أو تحويلات عليها، وإنما يستخدم المنطق المألوف.

ومثل هذه الاختبارات تناسب متطلبات النظم التـربوية التقليدية التى تهتم اهتمامًا أساسيًـا بالتفكير التقـاربى أو النمطى، ويمكن استـخدامها فى التنبـؤ بأداء الطلاب فى المجـالات الدراسية التى تعتـمد على قـدرات الذاكرة. غـير أن التوجـهات التـربوية المعـاصرة تؤكد أهمـية تنمـية استراتيـجيات التـفكير التـى تشجع الطلاب علـى إنتاج معلومات جديدة، وتوظيف إمكاناتهم العقلية فى مواقف ثرية متنوعة ، مما يتطلب بناء اختبارات من نوع مختلف لقياس التفكير التباعدى وتقييم نواتجه.

الابتكارية ومفهوم التفكير التباعدى :

تُعد الابتكارية Creativity من المفاهيم السيكولوجية التى نالت مزيدًا من اهتمام علماء النفس والتربية فى الآونة الأخيرة، غير أن هناك تباينًا ملحوظًا فى تحديد وتعريف مفـهوم الابتكارية نـظرًا لتعدد وجـهات النظر حـوله. إذ يرى بعض هؤلاء العـلماء أن الابتكارية تـتضمن عمليـة ممتدة عبـر الزمن وتتميـز بالأصالة Originality، والتواؤم Adaptiveness، والتحقيق Realization. وهذا يعنى أنـه ينبغى دراسة الأفـراد الذين تميزوا بمستوى مرتفع فى العـمل الابتكارى. ويرى بعض آخر أن الابتكارية هى القدرة على كشف شىء جديد، بينما يرى آخرون أن الابتـكارية ليست قدرة ، وإنما عملية أو عمليات سيـكولوجية تسهم فى تشكيل إنتـاج يتميز بالجدة والقيمـة، فى حين لا يركز البعض منهم فى مناقشته لمفهوم الابتكارية على العملية وإنما على الإنتاج ذاته.

أى أن هذه التـعريفات تـتراوح بين النظـر إلى الابتكارية عـلى أنها مـجرد حل للمشكلات يـتميز بعنصر الجـدة، والنظر إليها على أنها الـتحقيق الكامل والتعـبير عن جميع طاقات الفرد المتميزة.

ويرى ماكينون Mackinnon أنه ليس من المطلوب التـوصل إلى أفضل تعريف من بين هذه التعريفـات المتباينة نظرًا لأن الابتكارية تحمل جمـيع المعانى التى أشارت إليها وربمـا أكثر من ذلك. لذلك يفضل الـنظر إلى الابتكارية على أنها ظاهـرة متعددة الأوجه بدلاً من اعتبارها مفهومًا نظريًا يلزم تعريفه تعريفًا دقيقًا، مما يعنى أن الابتكارية Creativeness يمكن أن تضـم عددًا من الأوجه المتـرابطة. وأحد ميزات هذا الـتوجه وبخاصة فى دراسة الابتكارية من منظور متـسع، هو أنه يمكن تقسيم هذه الظاهرة التى تتسم بالتعقد إلى أوجه أو جوانب متمايزة يمـكن التعامل معها بدرجة أفضل من مفهوم الابتكارية بعامة.

وفى إطار هذا التوجُّه توجد أربعة أوجه للابتكارية هى: الابتكارية كإنتاج، والابتكارية كعملية، والشخص المبتكر، والموقف الابتكارى.

وهذه الأوجه الأربعة تتعلق بمجموعة من التساؤلات كالتالى:

(١) ما الإنتاج المبتكر ؟ وما الخصائص التى تُميزه عن غيره من الإنتاج المألوف؟

(٢) ما طبيعة العمليات الابتكارية؟ وما طبيعة ونوع العمليات السيكولوجية التى تسهم فى الإنتاج المبتكر؟

(٣) ما السمات والخصائص التى تميز الأشخاص المبتكرين عن غيرهم من الأشخاص الأقل ابتكاراً؟

(٤)ما خصائص الموقف الابتكارى والظروف الاجتماعية والثقافية ومناخ العمل الذى يُيسر ويشجع على استثارة الفكر والعمل الابتكارى؟

وعلى الرغم من أهمية دراسة هذه الأوجه المتعددة للابتكارية، إلا أن ما يعنينا هنا هو القدرات أو الوظائف العقلية المعرفية وما يتعلق بها من عمليات التفكير الابتكارى التى تسهم فى إمكانية التوصل إلى إنتاج ابتكارى.

ولعل الدراسات الثرية المتعلقة بالاستعدادات الابتكارية التى أجراها جيلفورد Guilford ومعاونوه فى مشروع بحث الاستعدادات Aptitude Research Project، بجامعة جنوب كاليفورنيا منذ عام ١٩٤٩ (Guilford; 1967; Guilford & Hoepfner , 1971) ، ودراسات تورانس ومعاونيه بجامعة مينسوتا (Torrance et al. , 1973) تعد من الدراسات الرائدة فى هذا المجال. وقد أسهمت دراسات وبحوث جيلفورد ومعاونوه فى اقتراح نموذج بنية العقل Structure of Intellect (SOI) الذى أوضحناه بالتفصيل فى الفصل التاسع، حيث أشرنا فيه إلى الأبعاد أو الأوجه الثلاثة المتعلقة بالمحتوى والعمليات والنواتج التى تم التوصل إليها باستخدام منهج التحليل العاملى.

ومن بين هذه العمليات الرئيسة التفكير التقاربى أو النمطى Convergent Thinking، والتفكير التباعدى أو المنطلق Divergent Thinking. وقد سبق أن أشرنا أن محدودية اختبارات الذكاء وأوجه قصور مفهوم نسبة الذكاء IQ فى النظر إلى القدرة العقلية أدى إلى زيادة الاهتمام بالتفكير التباعدى. وقد حاولت بعض الدراسات تعرُّف ما إذا كانت اختبارات هذا النمط من التفكير تقيس وظائف عقلية متفردة أم لا، أى ما إذا كان الذكاء العام يفسِّر الأداء فى كل من نمطى التفكير، أم أن هناك عاملاً خاصًا يمكن أن يفسِّر ارتباط اختبارات التفكير التباعدى فيما بينها، غير أن كرويلى

(Cropley , 1966) قد اقترح في دراساته أن أفضل حل وسط بين طرفي النقيض هو أن الوظائف العقلية المتعلقة بنمط التفكير التباعدي ترتبط ارتباطًا وثيقًا بالوظائف العقلية المتعلقة بنمط التفكير التقاربي. ودلل على ذلك بأن الأفراد الذين يستطيعون التوصل إلى حلول للمشكلات بتفكير منطلق يحصلون في غالبية الأحيان على درجات مرتفعة في اختبارات الذكاء التقليدية، ولكن أداؤهم في اختبارات التفكير التباعدي ربما يكون مرتفعًا أيضًا أو ربما لا يكون كذلك. بل إن بعض من يتميزون بتفكير تباعدي مرتفع ربما يكون أداؤهم في اختبارات الذكاء التقليدية منخفضًا. وقد استنتج كروبلي Cropley أن بعض الأفراد من ذوي القدرة العقلية المرتفعة ربما يفضّل توظيف أحد نمطي التفكير، بينما يفضّل بعضهم النمط الآخر.

قياس قدرات التفكير الابتكاري :

يُعد كل من جيلفورد Guilford وتوارنس Torrance من رواد حركة قياس قدرات التفكير الابتكاري، وقد أعدا اختبارات تقيس قدرات محددة تتعلق بهذا النمط من التفكير هي : الطلاقة Fluency ، والمرونة Flexibity، والأصالة Originality، وإدراك التفاصيل Elaboration، سوف نوضحها فيما يلي:

الطلاقة: هي القدرة على إنتاج عدد كبير من الأفكار والكلمات وأساليب التعبير عن الأشياء، أي طلاقة الأفكار Ideational Fluency، وطلاقة التداعي Associational Fluency، والطلاقة التعبيرية Expressional Fluency، فمثلاً يُقدم للفرد كلمة أو فكرة ويُطلب منه التفكير في أكبر عدد ممكن من الاستخدامات غير المألوفة للكلمة أو الشيء المعطى، أو مرادفاتها، أو تكوين جمل ذات خصائص معينة. وتقدّر نقطة واحدة لكل فكرة جديدة، فكلما زاد عدد الأفكار التي يُنتجها الفرد حصل على درجه أعلى في القدرة على الطلاقة.

المرونة: هي القدرة على إنتاج أفكار تُبين انتقال الفرد من مستوى تفكير إلى مستوى آخر، أو تحوّل تفكير الفرد بالنسبة لمهمة معينة، ويُقصد بذلك عدم اقتصار الفرد على خط فكري واحد، أي مرونة التفكير. ونظرًا لاختلاف محتوى المثير الذي يقدم للفرد ونوع ناتج التفكير فإن المرونة تعتمد على هذا المحتوى، فإذا كان المحتوى يتعلق بالمعاني اللغوية فإنه يُطلق عليها المرونة التلقائية Spontaneous Flexibility، وإذا كان يتعلق بالأشكال فإنه يطلق عليها المرونة التوافقية Adaptive Flexibility. وتقدر نقطة واحدة لكل تحوّل في التفكير، ويشير مجموع النقاط إلى القدرة على المرونة.

الأصالة: هي القدرة على إنتاج أفكار نادرة بالمفهوم الإحصائي للندرة، أي أفكار غير مألوفة وتباعدية ومميزة. وإحدى الطرق الشائعة الاستخدام في تقدير

الدرجات هي إعطاء نقاط للأفكار التي يُنتجها خمسة أفراد من بين مائة فرد، وكذلك تقدَّر الدرجات على ميزان متدرج يشتمل على خمس نقاط، حيث تُقدَّر نقطة واحدة للأفكار التي ربما تكون معروفة ، وتتدرج إلى خمس نقاط للأفكار التي تتميز بدرجة عالية من الجدة. والأفكار نفسها التي يُنتجها أكثر من خمسة أفراد من بين مائة فرد يقدر لها صفراً، ويشير المجموع الكلي للنقاط إلى الدرجة في هذه القدرة.

إدراك التفاصيل: هي القدرة على إضافة تفاصيل إلى فكرة رئيسة يُنتجها الفرد، ومثال ذلك أن يُقدَّم للفرد عدد من المربعات، ويُطلب منه أن يتوصل إلى أفكار لا يستطيع أحد غيره أن يُنتجها، وذلك برسم صور باستخدام المربع المعطى، كأن يرسم بابًا باستخدام المربع ويكتب عنوانًا للصورة. وإذا قرر فرد أن يضيف تفاصيل أخرى إلى صورة الباب، مثل بعض الزخرفات، أو فتحة للرؤية من الداخل، وهكذا، فإنه بذلك يضيف تفاصيلاً على الفكرة الرئيسة وهي الباب.

وتقدَّر نقطة واحدة لكل عنصر من العناصر التي أضيفت إلى الرسم، ويشير المجموع الكلي للنقاط إلى القدرة على إدراك التفاصيل.

وعلى الرغم من أن كلا من جيلفورد **Guilford** وتوارنس **Torrance** استندا إلى هذه القدرات الأربع في تقييم الابتكارية ، إلا أن جيلفورد حاول قياس كل منها بطريقة تتطلب من الفرد أداء مهام متعددة في ضوء المحتوى الشكلي بنوعيه، والرمزي، والسيمانتي، والسلوكي، ونواتج الوحدات، والفئات والعلاقات، والنظم، والتحويلات، والتضمينات، أي في ضوء ٥ × ٦ = ٣٠ قدرة من قدرات التفكير التباعدي كما سبق أن أوضحنا في شكل (٩ ـ ٤) .

ولكن حاول تورانس قياس القدرات الأربع الرئيسة نفسها بطريقة تتطلب من الفرد أداء عدد من المهام المعقدة التي تم تصميمها بحيث يوظف جميع هذه القدرات في آن واحد.

وقد أضاف تورانس مؤخراً عدداً آخر من القدرات، هي القدرة على التوليف **Synthesis**،والقدرة على التجريد **Abstraction** كما يتمثلان في إنتاج عناوين مميزة للرسوم التي ينتجها الفرد، ودمج شكلين أو أكثر في صورة متكاملة متعلقة بها، وكذلك القدرة على الغلق **Closure** المتمثلة في تأجيل إتمام مهمة معينة لمدة زمنية تسمح بالتأمل من أجل إمكانية التوصل إلى إنتاج أفكار تتميز بالأصالة . وفيما يلي توضيح لكل من هذين الاختبارين وغيرهما.

اختبارات التفكير الابتكاري :

(١) بطارية اختبارات الإنتاج التباعدي لجليفورد

Guilford Tests of Divergent Production :

أعد جليفورد هذه البطارية في إطار برنامجه البحثي ونموذج بنية العقل الذي سبق أن أشرنا إليه، وتقيس قدرات محددة للتفكير التباعدي، وتناسب بدرجة أفضل طلاب المرحلة الثانوية. ويمكن تصنيف اختبارات هذه البطارية على أنها اختبارات الابتكارية Creativity Tests. ومن أمثلة مفردات هذه الاختبارات: طلاقة الكلمات التي تتطلب سرعة كتابة قائمة بكلمات تشتمل على حرف معين، والطلاقة التعبيرية التي تتطلب تسمية أشياء أو موضوعات تنتمي إلى فئة معينة، وطلاقة التداعي التي تتطلب إنتاج مترادفات، والاستخدامات المتعددة التي تتطلب كتابة قائمة بما يحتمل أن يترتب على فعل معين، وتتطلب جميع هذه الاختبارات استجابات لفظية. غير أن البطارية تشتمل أيضًا على اختبارات تستخدم الأشكال، مثل اختبارات تكوين الأشياء الذي يتطلب رسم أشياء معينة استرشادًا بمجموعة محددة من التصميمات، واختبارات مشكلات تتعلق بأعواد الكبريت، وتعتمد على تقديم تصميم معين بأعواد الكبريت، ويطلب من الفرد تحريك عدد معين من الأعواد لإنتاج تصميم آخر.

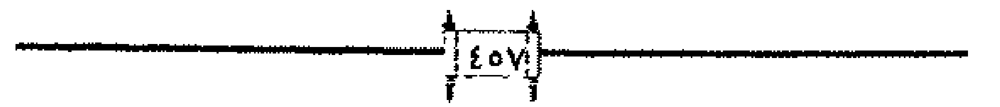
(٢) بطارية اختبارات تورانس للتفكير الابتكاري

Torrance Tests of Creative Thinking :

على الرغم من أن هذه البطارية تشبه بطارية جليفورد، إلا أن تورانس أعدها لتكون ملائمة للواقع التربوي الفعلي، ويُفترض أنها تصلح للتطبيق من الحضانة إلى المراحل الأعلى. وقد عرّف تورانس الابتكارية بأنها عملية استشعار الفجوات أو العناصر المفتقدة، وتكوين فروض تتعلق بها، واختبار صحة هذه الفروض، وتقرير النتائج ، وربما أيضًا إجراء تعديلات على الفروض وإعادة التحقق من صحتها (Torrance , 1962).

كما بيّن أنه يمكن تعريف الابتكارية من حيث الأصالة على أنها القدرة على التصور والابتعاد عن النمطية في الإدراك للتمكن من إعادة تكوين الصور الذهنية والأفكار والمشاعر في روابط تتميز بالجدة (Khatena &Torrance , 1973). وتنقسم البطارية إلى قسمين رئيسين هما: التفكير الابتكاري بالكلمات، والتفكير الابتكاري بالصور، وقد أضيف فيما بعد التفكير الابتكاري بالأصوات والكلمات.

وتم تصميم المفردات على شكل ألعاب أو صور. ويتضمن الجزء الخاص بالتفكير الابتكاري بالكلمات أنشطة متنوعة، كأن يُعطى الفرد رسمًا تخطيطيًا ويُطلب



منه كتابة قائمة بالتساؤلات التى يراها تتعلق بما يجرى فيما يمثله الرسم، أو تخمين الأسباب المحتملة لما يجرى، وكذلك تخمين ما يمكن أن يترتب على ما يجرى. ومن الأنشطة الأخرى اقتراح تحسينات يمكن إدخالها على لعبة، واقتراح استخدامات غير مألوفة لشيء مألوف، واقتراح تساؤلات حول هذا الشيء، وتخمين ما يترتب على فعل غير محتمل حدوثه.

أما الجزء الخاص بالتفكير الابتكارى بالصور فيشتمل على ثلاثة اختبارات تتعلق باستخدام شكل ملون فى تكوين صورة تتميز بالأصالة، واقتراح عنوان لها، ورسم أشياء باستخدام أزواج من الخطوط غير المنتظمة المعطاة، وإنشاء تصميمات باستخدام أزواج من الخطوط المتوازية أو الدوائر.

والجزء الخاص بالأصوات والكلمات يشتمل على اختبارين يُطَبَّقان باستخدام أجهزة تسجيل حيث تُقدم التعليمات والمثيرات ، ويتعلق أحد الاختبارين بالأصوات والصور، والآخر بالأصوات التى تحاكى الكلمات المسموعة والصور، ويستجيب الفرد لمفردات كل من الاختبارين لفظيًا بعد سماعه كل مثير صوتى حيث يعبِّر عما يقترحه أو يتصوره بشأن هذه المثيرات.

وقد فضَّل تورانس تسمية اختباراته «أنشطة» لكى يستبعد عوامل الخوف والقلق من الموقف الاختبارى.

ويتم تقدير الدرجات وفقًا للمحكات التى سبق أن أوضحناها. وتتكون البطارية من صيغتين زمن التطبيق الفعلى لأى منهما محدد وقدره ٤٥ دقيقة للجزء اللفظى، ٣٠ دقيقة للجزء غير اللفظى.. وقد وجد أن معامل ثبات تقديرات اختبارات كل من صيغتى البطارية تراوحت قيمته بين ٠,٥٠ ، ٠,٩٣ . ويشتمل دليل البطارية على أمثلة متنوعة تتعلق بكيفية تقدير درجات الاختبار والتدرب على ذلك، كما تشتمل على معايير الاختبار للفرق الدراسية المختلفة المستمدة من مختلف الأنظمة المدرسية.

وقد أجريت كثير من الدراسات حول هذه البطارية، وبرزت تساؤلات متعددة تتعلق بما تقيسه الاختبارات، وعلاقتها بالتحصيل الدراسى، والذكاء، ومحكات الإنجاز الابتكارى، مما يتطلب مزيدًا من البحث.

وعلى الرغم من ذلك فإن البطارية تعد جهدًا رائدًا فى مجال بناء وتقنين اختبارات الابتكارية.

(٣) اختبار التداعى الطليق Remote Associates Test :

أعد هذا الاختبار مدنيك Mednick، ومدنيك Mednick (١٩٦٧). ويقيس الطلاقة فى إنتاج تداعيات مألوفة لكنها صعبة الاستدعاء، وفحص هذه التداعيات، والانتقال

إلى تداعيات جديدة إلى أن يتوصل الفرد إلى أحد التداعيات المقبولة. فمثلا يُقدم للفرد مجموعات يشتمل كل منها على ثلاث كلمات مثل (مفاجأة، خط، عيد ميلاد) ويُطلب من الفرد التفكير في كلمة رابعة تُستخدم في الحياة اليومية وترتبط بكل مجموعة من هذه الكلمات، وتتطلب الإجابة التفكير في اتجاهات متعددة قبل أن يكتشف الفرد الكلمة التى فى ذهن مؤلفى الاختبار.

ويبدو أن الاختبار يتطلب استجابات مألوفة أى لا تتميز بالجدة، غير أنه يُفترض أن الأفراد الأكثر ابتكارية يكون انتاجهم للتداعيات المتعلقة بكلمات شائعة ضعيفًا، أما إذا كانت الكلمات غير شائعة، فإن انتاجهم للتداعيات يكون أقوى، أى أنه يمكنهم إنتاج تداعيات غير شائعة نسبيًا.

وقد أوضحت الدراسات المتعددة التى تناولت العلاقة بين الاختبار ومحكات مختلفة متعلقة بالابتكارية أن قيم معامل الارتباط تراوحت بين ٠,٥٥ ، ٠,٨٤ ، مما يشير إلى إمكانية الإفادة من الاختبار فى التنبؤ بالابتكارية.

ومن الجدير بالذكر أنه توجد اختبارات تقيس الابتكارية فى مجالات خاصة محددة، مثل اختبار القدرة الابتكارية AC Test of Creative Ability الذى أعدته شركة جنرال موتورز الأمريكية لقياس الابتكارية لدى المهندسين والمشرفين على العمل فى قسم كهرباء السيارات بالشركة، واختبار الابتكارية فى تصميم الآلات Owens Creativity Test for Machine Design الذى أعده أونز Owens عام ١٩٦٠، وهو اختبار أكثر صعوبة من الاختبار السابق، ومحتواه يتناول أمورًا فنية تتعلق بالآلات، ويُطبق على المهندسين. هذا بالإضافة إلى الاختبارات التى تقيس الابتكارية فى المجالات الفنية والموسيقية.

بعض المشكلات المتعلقة بقياس الابتكارية:

اتضح لنا مما سبق أن مجال قياس الابتكارية لا يزال فى مرحلة المهد إذا ما قورن بمجال قياس الذكاء وقدرات التفكير التقاربى حيث ترسخت كثيراً من الأسس والمبادئ التى يستند إليها المجال الأخير.

لذلك نجد أن قياس الابتكارية محاط بالعديد من المشكلات المنهجية والسيكومترية التى تستثير فكر علماء النفس والتربية وخبراء القياس فى محاولة التغلب عليها، وتطوير أساليب القياس فى هذا المجال الذى يزداد أهمية فى عالمنا المعاصر.

فالكشف عن الطاقات العقلية المبدعة واستثمارها لخير الإنسان ومواجهة التحديات فى عصر دائم التغير والانفتاح يعد ضرورة ملحة، وسوف نوضح فيما يلى بإيجاز بعض أهم هذه المشكلات:

مدى عمومية سمة الابتكارية:

تختلف اختبارات الابتكارية فيما بينها من حيث المحتوى، فبعضها يكون محتواه لفظيًا أو غير لفظي مثل الصور والأشكال المجردة والأصوات، وبعضها يجمع بين هذه المحتويات كما فى بطارية اختبارات تورانس. غير أن هناك اختبارات تقيس الابتكارية فى مجالات محددة كالهندسة والفنون، والموسيقى، والتأليف، والبحث العلمى، ويكون محتواها أكثر تقييدًا من الاختبارات التى تقيس الابتكارية بعامة.

وهذه الاختبارات النوعية لا يستطيع الإجابة عنها إلا المهتمون بمجالاتها ولديهم خبرة وتدرب في هذه المجالات. وهنا يبرز تساؤل حول ما إذا كانت الابتكارية سمة نوعية بدرجة كبيرة أم أنها سمة عامة. فإذا كانت هذه السمة نوعية، فإن المهندس أو الفنان أو المؤلف المبتكر ربما لا يكون كذلك فى مجالات أو مواقف تختلف عن مجال اهتمامه. أما إذا كانت السمة عامة فإن كلا من هؤلاء المبتكرين فى مجالات نوعية محددة ربما تظهر بعض علامات ابتكاريتهم فى مجالات أو مواقف مختلفة.

غير أن فاعليتهم فى ابتكار أفكار جديدة فى هذه المجالات الأخرى سوف تكون محدودة نظرًا لقلة تدريبهم فى هذه المجالات. ومع هذا فإن نزعتهم الابتكارية يمكن قياسها بالاختبارات نفسها التى تستخدم فى قياس الابتكارية فى المجالات الأخرى.

لذلك فإن المشكلة التى تُعد مجالا من مجالات البحث هى ما إذا كانت هناك بعض الخصائص المشتركة بين المبتكرين بغض النظر عن مجال اهتمامهم، وما إذا كانت هناك بعض الخصائص المتعلقة فقط بمجالات نوعية محددة.

علاقة اختبارات الابتكارية باختبارات الذكاء:

نالت هذه المشكلة الأساسية اهتمامًا كبيرًا من جانب علماء النفس والقياس، وقد أبرز ثورنديك (Thorndike , 1963)، ووالش وكوجان (Walsh & Kogan , 1965) وجيتزلس وجاكسون (Getzels & Jackson , 1962) هذه المشكلة فى التساؤل التالى: هل اختبارات الابتكارية والذكاء تقيس مجالات أو نطاقات معرفية مختلفة؟ فلقد بينت نتائج دراسات والش Walsh وكوجان Kogan أن الأدلة لا تؤكد الزعم بأن اختبارات الابتكارية الشائعة الاستخدام تقيس نطاقًا موحدًا من الوظائف المعرفية يختلف عن النطاق الذى تقيسه اختبارات الذكاء. وقد استند كل منهما فى استنتاجه إلى أدلة ثلاثة:

(أ) الارتباطات بين الاختبارات الفرعية التي تشتمل عليها مقاييس الذكاء كانت مرتفعة بعامة.

(ب) الارتباطات بين الاختبارات المختلفة التي تقيس الابتكارية كانت منخفضة بعامة.

(جـ) الارتباطات بين اختبارات الذكاء واختبارات الابتكارية كانت مساوية تقريبًا في مقدارها للارتباطات بين الاختبارات المختلفة للابتكارية ذاتها.

وقد أدى ذلك إلى استنتاج أن ما تقيسه اختبارات الابتكارية لا يعكس شيئًا أكثر مما تقيسه بالفعل اختبارات الذكاء، وأرجع الباحثان ذلك إلى قصور في بناء وتطبيق اختبارات الابتكارية.

ولمعالجة هذا القصور أجرى الباحثان دراسة على ١٥١ طفلاً من الصف الخامس الابتدائي، واعتمدا في تقييم السلوك الابتكاري على أسلوب يسمح للأطفال بوقت كاف للإجابة عن مفردات الاختبار التي قُدِّمت على شكل ألعاب بتعليمات تُقلل من التنافس فيما بينهم، وبذلك تصبح الظروف مناسبة لما يتطلبه السلوك الابتكاري. وقد وجدا أن متوسط قيم الارتباطات بين درجات كل من اختبارات الابتكارية والذكاء مساويًا ٠,٠٩، ومتوسط قيم الارتباطات بين الاختبارات المختلفة للابتكارية ٠,٤١، وبين الاختبارات المختلفة للذكاء ٠,٥١. وهذا ربما يكون دليلاً على أن هناك بُعدين مستقلين للنطاق المعرفي الذي تقيسه كل من اختبارات الابتكارية واختبارات الذكاء.

لذلك اقترح الباحثان ضرورة الاعتماد في بناء اختبارات الابتكارية على نموذج يختلف عن نموذج بناء اختبارات الذكاء، أي يعتمد على فعل الابتكار ذاته كما يحدث بعيدًا عن المواقف الاختبارية.

ويؤخذ على هذه الدراسة وغيرها من الدراسات التي تناولت العلاقة بين اختبارات الابتكارية واختبارات الذكاء اعتمادها على عينات من الأفراد تتميز بذكاء يفوق المتوسط مما يجعل مدى الدرجات ضيقًا، وبذلك تنخفض قيمة معامل الارتباط، أما إذا اشتملت العينة على مدى متسع للذكاء فسوف يكون هناك ارتباط بين نوعي الاختبارات. غير أن قيمة الارتباط تتباين تباينًا ملحوظًا وفقًا للاختبارات المستخدمة، وطريقة تطبيقها، وطبيعة عينة الأفراد المستخدمة.

ففي إحدى الدراسات طبق كروبلي (Cropley , 1966) اختبارات ابتكارية واختبارات ذكاء على عينة عشوائية من الأطفال، ووجد أن الارتباط بين درجات نوعي الاختبارات ٠,٥١، مما جعله يستنتج أن القدرات المتعلقة بكل من التفكير التقاربي والتفكير التباعدي متمايزة عامليا، ولكنها ليست مستقلة.

وقد توصلت دراسة جيتزلز Getzels ، وجاكسون Jackson إلى أن الذكاء فوق المتوسط ضرورى للابتكارية، ولكن العكس غير صحيح، أى أن الذكاء المرتفع نسبيًا ضرورى للابتكارية ، ولكنه ليس كافيًا.

وعمومًا، فإن هذه المشكلة تُعد ميدانًا خصبًا لمزيد من البحوث والدراسات استنادًا إلى منهجيات وأساليب إحصائية مختلفة.

صدق اختبارات الابتكارية :

تُعد مشكلة صدق اختبارات الابتكارية من المشكلات الأساسية أيضًا، حيث إن رصيد المعلومات المتوافر حول هذه المشكلة محدودًا، فالدراسات التى تناولت السمات السيكولوجية المتعلقة بالابتكارية، والأساليب المناسبة للتنبؤ بها فى مواقف الحياة الفعلية كانت نتائجها متعارضة.

فعند تقدير صدق اختبارات الابتكارية تُستخدم في معظم الأحيان تقديرات المعلمين كمحك على الرغم مما يشوبه من عيوب. فهذا المحك ربما يكون أكثر ملاءمة في مجال التحصيل الدراسى أو السلوك القيادى، ولكنه لا يُناسب مجال الابتكارية، وكذلك تُستخدم أحكام الأقران كمحك لتقدير صدق هذه الاختبارات على الرغم من تأثر هذه الأحكام بمتغيرات متعددة مما يجعلها قليلة الفائدة (Torrance,1959).

والحقيقة أن تحديد محكات لتقدير صدق اختبارات الابتكارية أكثر صعوبة من تحديد محكات لتقدير صدق اختبارات الذكاء. فالإنتاج الابتكارى يتميز بالندرة ولا يسمح بسهولة تصنيفه على ميزان منتظم، ومع هذا فإن الدراسات التى استخدمت اختبارات التفكير التباعدى توصلت إلى نتائج متباينة ولكنها تشير إلى انخفاض قيم معامل الصدق التنبؤى لهذه الاختبارات (Wallach & Wing 1969).

وعلى الرغم من أن الأساليب التى اتبعها جيلفورد فى بناء بطارية اختبارات الابتكارية تكفل الصدق العاملي والصدق التمايزى لهذه البطارية، إلا أن الارتباط بين اختباراته ومحكات مرتبطة بمواقف الحياة الفعلية كان منخفضًا. كما أن نمط الارتباط بين درجات اختبارات تورنس ومحكات خارجية متنوعة لم يكن واضحًا.

لذلك فإن بطاريات اختبارات الابتكارية المتوافرة يمكن استخدامها لأغراض البحث، ولكنها لم تصل فى تطورها إلى مرحلة تسمح بتفسير درجات اختباراتها تفسيرًا يمكن الإفادة منه فى الواقع الفعلى، مما يتطلب إجراء دراسات صدق شاملة ومنظمة لتوضيح العلاقة بين اختبارات الابتكارية، ومقاييس القدرات العقلية الأخرى، ومحكات خارجية واقعية.



ثبات درجات اختبارات الابتكارية:

بيّنت كثير من الدراسات أن ثبات درجات اختبارات الابتكارية منخفضاً بدرجة واضحة، ولعل هذا يرجع إلى الصعوبات المتعلقة بتقدير درجات هذه الاختبارات، حيث إن مفرداتها تتطلب إجابات مفتوحة في وقت محدد. ومما يزيد من حدة هذه المشكلة انخفاض الصدق نظراً لصعوبة التحقق مما إذا كانت هذه الاختبارات تقيس الابتكارية بالفعل، وذلك لأن ما يُطلق عليه ابتكاري يختلف اختلافاً كبيراً من شخص إلى آخر، ومن ثقافة إلى أخرى. وهذا يدعو إلى أهمية التوصل إلى أساليب لتقدير درجات مفردات هذه الاختبارات تساعد في الحكم الموضوعي على مدى تميز الإجابات بالجدة، ولعل مفهوم الأصالة يُعد مدخلاً لتحقيق ذلك.

فلكي نحافظ على ميزات المفردات التي تتطلب إجابة مفتوحة وفي الوقت نفسه يمكن تصحيح الاختبار بالطرق الآلية، يمكن أن تكون الاستجابة لفظية، ويُفضل أن تكون كلمة واحدة.

فمثلاً يُقدم للفرد فقرات تتعلق بمجال معين يحدد في نهاية كل منها المطلوب منه، ويلي كل فقرة مفردة اختيار من متعدد تتضمن عدداً من البدائل التي يشتمل كل منها على بعض الحروف المتفرقة يفصل بينها عدد محدد من المسافات ، مثل (هـ ـ ـ ل م ـ ـ أ) واحد هذه البدائل التي تبدأ بحرف معين يكون صحيحاً من وجهة نظر مؤلف الاختبار.

وقد استخدم فلانجان (Flanagan , 1963) هذا الأسلوب في أحد اختباراته ، كما استخدمته بعض الاختبارات الأخرى. غير أنه يُعاب على هذا النوع من المفردات أن الفرد الذي يتسم بمهارة في استخدام الكلمات ربما يتوصل إلى الحل الصحيح، دون تفكير منطلق ، وهذا بالطبع ليس الهدف من الاختبار. كما أن الفرد المبتكر ربما يفكر في بديل متميز يختلف عن البدائل المعطاة مما يجعل الفرد يشعر بأن الاختبار يحد من تفكيره المنطلق. ولعل اختبار التداعي الطليق الذي سبق أن أشرنا إليه يُعد من هذا النوع إلى حد ما.

ومن هذا يتضح أن قياس وتقييم الابتكارية بقدر كبير من الدقة والمهارة كما هو مُتبع في قياس الذكاء يتطلب مزيداً من الدراسات الإمبيريقية، والسيكومترية من منظور علم النفس المعرفي من أجل التعرف على العمليات المعرفية التي تُسهم في الإنتاج الابتكاري وما يرتبط بذلك من متغيرات غير معرفية، مثل الدافعية، والخصائص الشخصية،والبيئة المنزلية، والمدرسية، وغير ذلك من المتغيرات التي يمكن أن تلقى مزيداً من الضوء على ظاهرة الابتكارية.

خلاصة

إن اختبارات الاستعدادات الخاصة تعد مقاييس معرفية تُستخدم في التنبؤ بنجاح الفرد قبل التحاقه ببرنامج تعليمي أو تدريبي معين، كما تُستخدم في أغراض الانتقاء والتسكين والإرشاد والتوجية التربوي والمهني. وتشتمل الاستعدادات على تجمعات من القدرات الخاصة المتعلقة بها ويمكن قياسها. وتتعدد أنواع هذه الاختبارات نظرًا لتعدد وتنوع الاستعدادات الخاصة.

فهناك بطاريات متعددة العوامل تشتمل على اختبارات فرعية منفصلة لقياس القدرات المتعلقة بعدد من الاستعدادات الأكاديمية أو المهنية التي تهدف البطارية لقياسها. ويمكن رسم صفحة نفسية أو بروفيل لدرجات الفرد في الاختبارات الفرعية لتوضيح جوانب قوته وضعفه، ولكن ينبغي الحيطة عند تفسير دلالة الفروق بين الدرجات في هذه الاختبارات.

وهذا يعتمد على ثبات درجات الاختبارات، ودرجة ارتباط كل منها بالآخر، حيث ينبغي أن يكون الثبات مرتفعًا، والارتباط فيما بينها منخفضًا وهناك اختبارات استعدادات نوعية يقيس كل منها استعدادًا خاصًا واحدًا، مثل الاستعداد الميكانيكي، والاستعدادات النفسية الحركية، والاستعداد الكتابي، والاستعداد الموسيقي، والاستعدادات الفنية، والابتكارية، واستعدادات أخرى تُستخدم لأغراض خاصة يتطلب بناؤها أساليب منهجية منظمة لتحليل العمل.

ويستند قياس الاستعدادات الخاصة إلى مجموعة من الافتراضات الأساسية التي ينبغي مراعاتها. وتختلف أساليب بناء اختبارات الاستعدادات الخاصة باختلاف نوع الاختبار وتكوينه والغرض من استخدامه. وينبغي أن تتميز جميع هذه الاختبارات بصدق تنبؤي مرتفع. وللتحقق من ذلك يجب العناية باختيار المحك المناسب الذي تتوافر فيه الشروط المطلوبة، وكذلك تحديد درجة أو درجات القطع استنادًا إلى الأساليب المنطقية، أو الإمبيريقية. غير أن اختبارات الابتكارية تختلف عن غيرها من الاختبارات في أسلوب بناء مفرداتها وتقدير درجاتها، وصعوبة تحديد محك مناسب لتقدير صدقها، كما أن ثبات درجاتها يكون منخفضا نسبيًا؛ لذلك فإن هذا النوع من الاختبارات يتطلب مزيدًا من البحث والتطوير بفكر ابتكاري.



تطبيقات القياس والتقويم التربوى والنفسى

(ثانيا)قياس وتقويم الجوانب غير المعرفية (الوجدانية والشخصية)



الفصل الحادى عشر

قيـــاس الميـول

* مقدمة
* قياس الميول من منظور تاريخي
* الميول وعلاقتها بالاستعدادات الخاصة
* الأنماط المختلفة للميول
* أساليب بناء استبيانات الميول
* بعض المشكلات المتعلقة بقياس الجوانب غير المعرفية
* بعض المشكلات المتعلقة بالاستبيانات المقننة للميول
* مشكلات تفسير درجات استبيانات الميول
* أساسيات بناء استبيانات الميول المهنية
* قياس الميول باستخدام الاستبيانات المقننة
* تعقيب عام على استبيانات الميول
* استخدامات استبيانات الميول
* تقويم ميول الطلبة فى الصف الدراسى
* أدوات قياس وتقويم ميول الطلبة فى الصف الدراسى

مقدمة :

انصبَّ اهتمامنا فى الباب الثانى على أساليب وأدوات قياس وتقويم الجوانب المعرفية للشخصية المتعلقة بالتحصيل الأكاديمى والكفايات والذكاء والاستعدادات الخاصة متضمنة الابتكارية ، وجميعها تهتم بقياس الأداء الأقصى للفرد Maximum Performance، أى أنها تُوضح ما يستطيع الفرد إنجازه فى ظروف معينة، وتُعد منبئات جيدة بالنجاح الأكاديمى والمهنى.

غير أن هناك جوانب أخرى للسلوك الإنسانى لا تقل أهمية عن الجوانب المعرفية، وينبغى قياسها وتقويمها والإفادة من نتائجها فى اتخاذ القرارات المتعلقة بالأفراد، وتتعلق بالجوانب غير المعرفية أو الوجدانية أو الشخصية. وهذه الجوانب غير المعرفية Non - Cognitive لا تنفصل عن الجوانب المعرفية Cognitive نظراً لتكامل السلوك الإنسانى وتأثره بهذه الجوانب مجتمعة، فإذا اقتصرنا على قياس قدرات واستعدادات الفرد العقلية أو المعرفية فإن هذا لا يعنى أن الجوانب المعرفية تؤثر وحدها فى أدائه، وكذلك إذا اقتصرنا على قياس الجوانب غير المعرفية ، فإن هذا لا يعنى أن أداء الفرد لا يتأثر بالجوانب المعرفية، فأثر كل من هذه الجوانب ينعكس فى أداء الفرد أو سلوكه لأنها تمثل أوجه متعددة للشخصية الإنسانية تتفاعل فيما بينها. وتهتم أدوات قياس الجوانب غير المعرفية بتقييم الأداء المميز للفرد Typical Perfomance.

وعلى الرغم من أن إسهام هذه الأدوات فى التنبؤ بالنجاح الأكاديمى والمهنى أقل مما تسهم به أدوات قياس الجانب المعرفى، إلا أنها تفيد بدرجة أكبر فى الإرشاد والتوجيه، حيث تضيف معلومات مهمة عن الفرد إلى جانب ما تقدمه الأدوات السابقة. وتتعدد الخصائص والمتغيرات التى تتضمنها الجوانب غير المعرفية أو الوجدانية Affective، حيث تشمل الانفعالات والعواطف، والميول، والاتجاهات، والمعتقدات، والقيم، والدوافع، والسمات المزاجية، والحس الاجتماعى، وغير ذلك من سمات الشخصية الإنسانية.

غير أننا سوف نركز فى هذا الباب على بعض هذه المتغيرات المهمة وهى الميول، والاتجاهات، والشخصية. وعلى الرغم من أن المفهوم العام للشخصية Personality يتضمن عادة جميع الخصائص أو المتغيرات السابقة، إلا أننا سوف نميز الميول والاتجاهات عن سمات الشخصية والسمات المزاجية ، وقد حظت هذه المتغيرات باهتمام كبير من جانب علماء النفس والتربية، وتناولتها كثير من الدراسات

بالبحث والتطوير. وسوف نقتصر في هذا الفصل على مناقشة مفهوم الميول وما يتعلق به من أدوات قياس، وكيفية بناء هذه الأدوات واستخداماتها والمشكلات الخاصة بها، ونتناول الاتجاهات، والشخصية في الفصلين التاليين.

قياس الميول من منظور تاريخي :

لعل أول محاولة للبحث في مجال الميول هي الدراسة التي أجراها ثورنديك E. Thorndike ونُشرت في المجلة الشهرية للعلوم العامة Popular Science Monthly عام ١٩١٢ وتتعلق باستقرار الميول وعلاقتها بالقدرات. واعتمدت هذه الدراسة على عينة تشتمل على ١٠٠ طالب بالكليات حيث طُلب من كل منهم ترتيب ميوله كما يتذكرها منذ كان في المرحلة الابتدائية، والثانوية، وعند التحاقه بالكلية، وكذلك ترتيب قدرته في هذه المراحل، وأوجد الارتباطات بين مختلف التراتيب، واستنتج أهمية الميول المبكرة وعلاقتها الواضحة بالقدرة مما يبين أن كلا منها يدل على الآخر، كما استنتج أن هذه العلاقة تتميز بالتعقد. والمحاولة المبكرة الثانية تتعلق باستبيان أعده كيلي Kelley عام ١٩١٤ اشتمل على أسئلة تستوضح تفضيلات الأفراد الترويحية، غير أن هذا الاستبيان لم ينل حقه من اهتمام الباحثين.

وقد بدأت النشأة الفعلية لاستبيانات الميول كما نعرفها الآن في معهد كارنجا التقني بولاية بيتسبرج الأمريكية بعد الحرب العالمية الأولى بمدة قصيرة . فقد عقد هذا المعهد حلقة بحث تتعلق بالميول ، وقام الأساتذة وطلاب الدراسات العليا ببناء ما يقرب من ألف فقرة للتمييز بين ميول العاملين في المهن المختلفة، وأسهمت هذه الحلقة إسهامًا كبيرًا في تشجيع الباحثين على بناء استبيانات مقننة للميول على أسس علمية منظمة، وأُعدت بالفعل كثيرًا من هذه الاستبيانات. فقد قامت مجموعة من علماء النفس الصناعي بريادة بنجهام Bingham بدراسة المشكلات النفسية التطبيقية بما في ذلك مشكلة قياس الميول.

وفي عام ١٩٣١ أصدر فراير Fryer كتابًا بعنوان «قياس الميول» لخص فيه التطورات الماضية التي حدثت في هذا المجال ، حيث أشار فيه إلى محاولة مور Moore عام ١٩٢١ قياس الميول الميكانيكية والاجتماعية للمهندسين، ومحاولة ريم Ream عام ١٩٢٤ التمييز بين الناجحين وغير الناجحين من العاملين بالمهن التجارية في ضوء ميولهم، وبحوث فرايد Freyd عام ١٩٢٤ التي أجريت على مقاييس مور Moore في محاولة التمييز بين الميول الميكانيكية والاجتماعية لمجموعات من عامة الناس.

وقد شهد العقد الثانى من القرن العشرين بداية بناء الاستبيانات المقننة للميول بإشراف سترونج Strong الذى وافق تيرمان Terman على تعيينه أستاذًا بجامعة ستانفورد عام ١٩٢٤ ، وكان تيرمان يجرى بحوثه فى ذلك الوقت على الطلاب من ذوى الذكاء المرتفع. ونظرًا لتأثر سترونج بنتائج حلقة البحث التى أشرنا إليها فقد قام بإعداد استبيان للميول المهنية عام ١٩٢٦ ، وتمت طبعته الأولى عام ١٩٢٧ ، ودليله ووثائقه المرجعية عام ١٩٢٨ ، وصدرت أول معلومات عن تصحيح هذا الاستبيان آليًا عام ١٩٣٠ . ونظرًا لانتشار استخدام هذا الاستبيان، فقد تناولته العديد من البحوث والدراسات الطولية، وأجريت عليه عدة تعديلات لتوسيع نطاقه خلال العقود التالية، وصدرت طبعته الحديثة التى يتم تصحيحها باستخدام الحاسوب عام ١٩٧٤ .

ومن التطورات التاريخية الأخرى المهمة المتعلقة بقياس الميول الاستبيان الذى أعده كيودر Kuder عام ١٩٣٩والذى يُعرف بسجل كيودر للتفضيل Kuder Preference Record، ويختلف فى أسلوب ومنهجية بنائه عن استبيان سترونج، وكذلك البحوث المتعلقة بالقياس الموضوعى للميول التى أجراها علماء النفس بسلاح الطيران الأمريكى أثناء الحرب العالمية الثانية.

وفى عام ١٩٤٣ أعد ثورب Thorpe استبيان الميول المهنية Occupational Interest Inventory ، ويختلف أيضًا عن سابقيه فى أسلوب ومنهجية بنائه. وفى عام ١٩٤٨ أعد جيلفورد Guilford ومعاونوه استبيان مسح الميول Guilford's Interest Survey استندوا فى بنائه على منهج التحليل العاملى، ويمكن باستخدامه الحصول على درجات منفصلة للميول المهنية والميول غير المهنية.

ونظرًا لاعتماد معظم مقاييس الميول على الجانب اللفظى ، فقد أعد جايست Geist استبيان الصور Geist Picture Inventory ، غير أنه توجد أدلة محدودة عن صدق مثل هذه الاستبيانات غير اللفظية. وقد ازداد بعد ذلك الاهتمام بقياس الميول غير المهنية أى المتعلقة بالأنشطة الحياتية المختلفة، مثل استبيان كيودر لمسح الميول بعامة الذى نشر عام ١٩٦٤ ، وكذلك بناء استبيانات لخريجى المدارس الثانوية الذين لا يستمرون فى الدراسة وإنما يلتحقون بالعمل ، ومثال ذلك استبيان مينيسوتا للميول المهنية. وتتوالى الدراسات والبحوث حتى وقتنا الحاضر فى هذا المجال، وبخاصة علاقة الميول بجوانب السلوك المختلفة المعرفية وغير المعرفية.

الميول وعلاقتها بالاستعدادات الخاصة :

لعل من المناسب فى مستهل هذا الفصل أن نوضح المقصود بالميول وعلاقتها بالاستعدادات الخاصة، وما إذا كانت تفيد فى التنبؤ بهذه الاستعدادات أو التنبؤ بالنجاح فى العمل، فالميل يمكن اعتباره تهيؤاً أو قابلية للمشاركة فى نشاط معين، وبتحديد أكثر فإن الميل شرط إدراكى Perceptual Condition يتضمن الوعى المعرفى والشعور الوجدانى معًا فى نمط معين من المشاعر. ومن الممكن أن تكون هذه المشاعر إما مؤقتة أو دائمة، وهذا يعتمد على كم وكيف الخبرة التى أدت إلى هذه المشاعر، وعادة يشار إلى الميل بأنه قابلية الفرد لأن ينتقى انتقاء حرًا أنشطة محددة تستثير لديه مشاعر سارة مثل القراءة فى المجالات العلمية، أو تنسيق الزهور، أو ممارسة لعبة رياضية معينة، أو تفضيل العمل فى مجالات مهنية محددة.

ومصطلح «الميل» يُعد نوعًا من المتغيرات المتصلة التى ربما تتغير فيه المشاعر تغيرًا منتظمًا على حسب الظاهرة أو الشىء المرتبطة به، غير أنه يمكن التمييز بين التعبير اللفظى للفرد عن ميله تجاه نشاط معين، ومشاركته الفعلية فى هذا النشاط. إذ يمكن أن يُعبر الفرد عن عدم ميله للموسيقى، ومع هذا ربما يقوم عادة بالعزف الجيد على إحدى الآلات الموسيقية، وبالمثل فى كثير من الأنشطة. لذلك ينبغى أن يكشف الفرد عن ميوله فى نشاط معين يقوم به، فالميل يتطلب نشاطًا معينًا ومشاركة فى هذا النشاط.

فالفرد الذى يميل إلى إصلاح الآلات الميكانيكية يُعد لديه ميل من مستوى معين، ربما المستوى المهارى. أما إذا كان الفرد لا يميل إلى التعامل المباشر مع هذه الآلات، وإنما يميل إلى التصميمات المجردة المتعلقة بها، فإن ميله يكون من مستوى آخر، ربما المستوى المهنى المتخصص. لذلك فإن اعتبار الميل فى كل من الحالتين يمثل الميل الميكانيكى ربما يُحدث بعض الخلط.

فعدم وضوح تعريف هذا الميل له انعكاساته على ما إذا كانت توجد علاقة بين الميل والقدرة أم لا توجد. وعمومًا ينبغى أن يوصف الميل بعبارات أنشطة محددة، كأن نقول أن شخصًا يميل إلى التحدث عن السيارات، ويقوم بإصلاحها، والاطلاع على مستحدثاتها، أو يميل إلى رسم تصميمات هندسية تتعلق بها. لذلك فإن كثيرًا من تعريفات الميول تُركز على الأنشطة التى يُفضلها الفرد أو لا يُفضلها أو لديه نزعة للقيام بأنشطة محددة، أو يحاول تجنبها. ومن الناحية الإجرائية يشار إلى الميول بأنها مهنية أو تربوية، على الرغم من أن الفقرات أو العبارات التى تشير إلى الأنشطة المحددة نفسها يمكن أن تستخدم فى كل من الحالتين.

وعادة يتم التمييز بين الميول والقدرات المستثمرة، أى التحصيل الدراسى والذكاء والاستعدادات الخاصة التى تشير إلى مهارات أو كفايات معينة. ويتضح هذا التمييز إذا أردنا تقييم مهارات أو كفايات الطلاب فى العلوم مثلاً فى مقابل تقييم تفضيلاتهم أو مدى إقبالهم على إجراء التجارب العلمية.

ومفهوم الميل يماثل مفهوم الذكاء أو القدرات فى أنه تجريد للسلوك Abstraction، أى أن مجمل الأنشطة التى يشارك فيها الفرد بمحض اختياره والتى يبدو أنه يجمعها شيء مشترك، مثل تفضيل الفرد للعمل الميكانيكى، يمكن تفسيرها بأنها تُعبر عن ميل الفرد للقيام بأعمال أو أنشطة ميكانيكية. غير أنه قد تبين بعامة أن علاقة الميول بالقدرات العقلية منخفضة، إذ يمكن أن يكون لدى الفرد ميل مرتفع نحو عمل معين، ولكن أداؤه المتعلق بهذا العمل يكون عاديًا، أو ربما تكون قدرته مرتفعة على أداء العمل ولكن لا يكون لديه ميل نحوه.

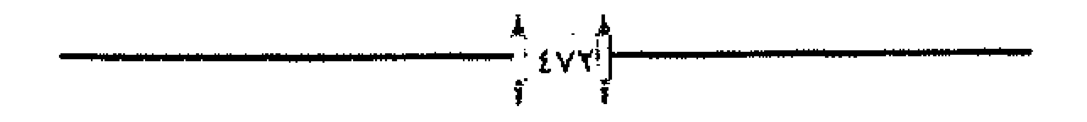
فالميول تشير إلى نطاق من التفضيلات مثل الطب، أو الهندسة، أو الأعمال التجارية، بينما تشير القدرات إلى استطاعة الفرد أداء مستوى معين من مستويات مهارة أو إنجاز معين، إذ ربما يستطيع أن يكون متخصصًا فى مهنة معينة أو ماهرًا أو شبه ماهر فى المهنة نفسها.

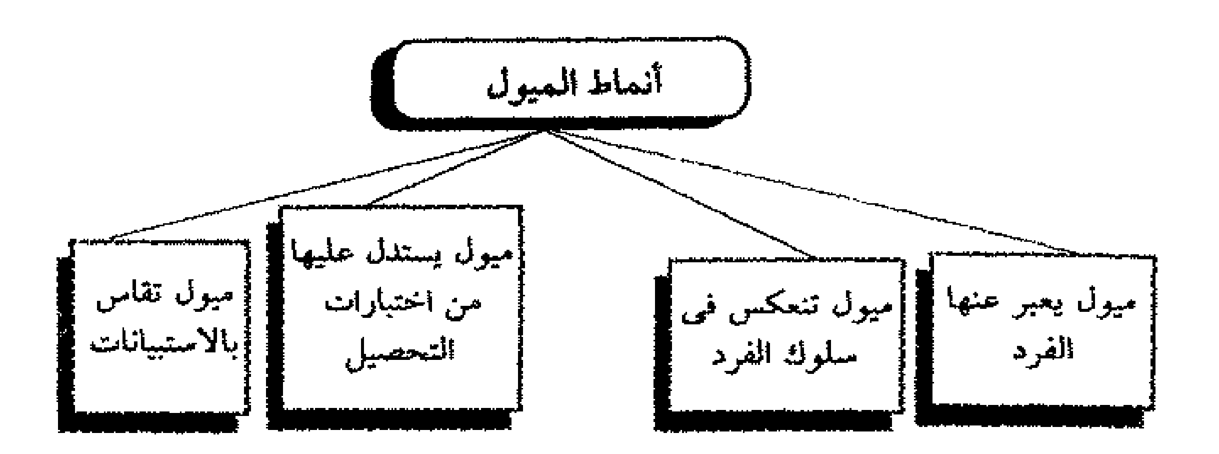
ومع هذا فإن الميول تتفاعل بلا شك مع القدرات، فالقدرات ليست ثابتة دومًا، وإنما تتأثر بالتعلم المستمر والخبرة وتغير متطلبات المهن والأعمال. فالفرد يمكن أن يغير مهنته أو تخصصه داخل المهنة الواحدة. غير أن دراسات سترونج (Strong, 1955)، وكامبل (Campbell, 1971) أوضحت أنه فى إطار الأقسام العريضة للمهن مثل الطب، والهندسة، والأعمال التجارية، يكون هناك استقرار نسبى للميول فى مرحلة الشباب، ولا ينزع معظمهم إلى تغيير مهنته تغييرًا ملحوظًا بقية حياته.

وعموما فإن اتخاذ القرارات التربوية والمهنية المتعلقة بالفرد يتطلب الاستناد إلى معلومات متعددة من مصادر مختلفة.

الأنماط المختلفة للميول:

يمكن تصنيف الميول فى أربع مجموعات استنادا إلى الأسلوب الذى يُتبع فى الحصول على المعلومات، وقد حدد سوبر وكرايتس (Super & Crites, 1962) هذه المجموعات التى يوضحها الشكل التخطيطى (١١ ـ ١) التالى:





شكل (١١ - ١) يوضح الأنماط المختلفة للميول

وفيما يلى توضيح لكل من هذه الأنماط:

(١) الميول التى يُعبر عنها الفرد: يمكن تحديد هذه الميول بأن يُطلب من الفرد التعبير شفاهة أو كتابة عن الأنشطة المهنية أو غير المهنية التى يُفضلها عن غيرها ويستمتع أو لا يستمتع بها. وعلى الرغم من أن هذه الطريقة تُعد أبسط طرق الحصول على معلومات تتعلق بميول الفرد، إلا أنه يُعاب عليها أنها لا تتميز بالثبات وتتأثر بخلفية الفرد وخبراته وخيالاته، وما يمكن أن يكون مقبولاً اجتماعيًا، كأن يُعبر عن ميله للموسيقى الكلاسيكية أو العمل فى مجال الفضاء. كما أنه ربما يحاول عدم التعبير عن ميله نحو الأعمال الحرفية، ولكنه يوضح ميله نحو الهندسة بدلاً عنها على الرغم من نفوره من الرياضيات.

ومع هذا فإن المعلمين يستندون إلى الميول التى يُعبر بها الطلاب عن أنفسهم Expressed Interests فى معاونة الطلاب اختيار موضوعات القراءة والتعبير، والوسائط التعليمية المناسبة، واستثارة المناقشات الجماعية.

(٢) الميول التى تنعكس فى سلوك الفرد: فالميول تتكشف من خلال الأنشطة الحرة التى يقبل الفرد عليها ويمارسها أو يشارك فيها Manifested Interests فالفرد الذى يسهم فى العمل التطوعى، والطالب الذى يشارك بانتظام فى الجمعية العلمية بالمدرسة يعكس بذلك ميله نحو هذه الأنشطة.

ويمكن تعرُّف ذلك عن طريق الملاحظة المباشرة للفرد أو محاولة معرفة هواياته وأنشطته التى يمارسها بالفعل، فالفرد الذى ينسق الزهور بانتظام فى حديقة منزله، أو يقتنى التحف، أو يقوم بإصلاح الأعطال فى أجهزته المنزلية، فإنه يكشف بذلك عن

ميله نحو نشاط معين. غير أن هذه الميول ربما تكون انتقالية، أى ربما يتحول اهتمام الفرد من نشاط إلى آخر لأسباب متعددة. فالفرد الذى يذهب إلى النادى ربما يجعلنا نظن أن لديه ميل للمشاركة فى الألعاب الرياضية، فى حين أنه يرى أن هذا المجال مناسب لتكوين علاقات اجتماعية، أو أنه يذهب لمجرد قضاء وقت الفراغ.

فالميول التى يعبِّر عنها الفرد لا تطابق بالضرورة الميول التى تنعكس فى سلوكه، ولكن كثيراً من الأفراد الذين يشاركون بانتظام فى أنشطة حرة يعبِّرون بالفعل عن ميلهم الإيجابى نحو هذه الأنشطة.

(٣) الميول التى يُستدل عليها من اختبارات التحصيل : يمكن الاستدلال على ميول الطالب من درجاته فى اختبارات التحصيل فى المجالات الدراسية المختلفة Inferred Interests. وهنا يُفترض أن الدرجات المرتفعة فى أحد هذه المجالات تتوافق مع ميل الطالب نحو هذا المجال.

غير أن كثيراً من الباحثين - كما سبق أن ذكرنا - وجدوا أن الارتباط بين الميول المقاسة والقدرات منخفضة. ومع هذا فإنه إذا قارنا ترتيب درجات الطالب فى استبيان للميول بدرجاته فى اختبارات التحصيل المتعلقة بها، فإنه ربما نلاحظ أن هناك تطابقًا بين درجاته المرتفعة فى مجال دراسى معين وميله المرتفع نحو هذا المجال.

(٤) الميول التى تقاس بالاستبيانات : يمكن تحديد ميول الفرد عن طريق قياسها باستبيانات مقننة حيث يُقارن بين ميوله فى أنشطة متعددة Inventorized Interests. وهذه الاستبيانات تتطلب من الفرد اختيار النشاط الذى يفضله أو لا يفضله من بين مجموعة كبيرة من الأنشطة المتعلقة بالمهن والمجالات الدراسية والترويحية والأنشطة المصاحبة والهوايات.

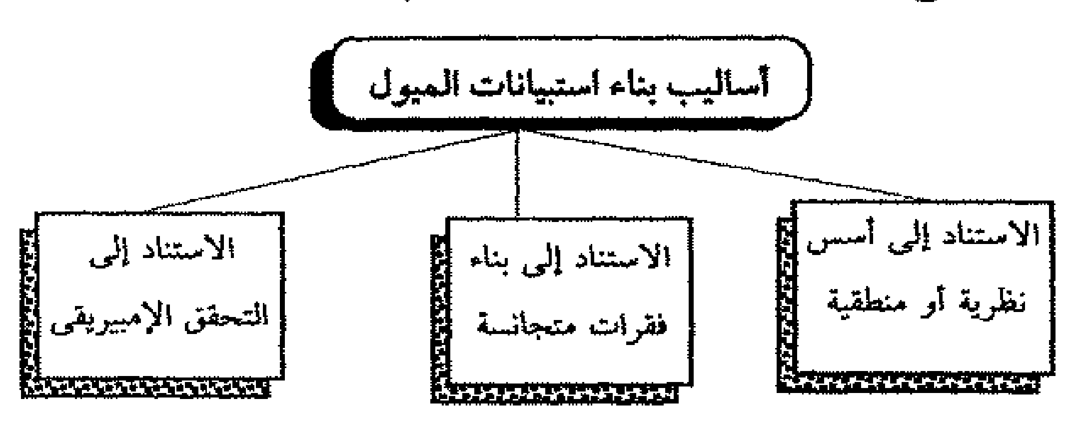
كما تتطلب منه ترتيب ظروف العمل المختلفة من حيث أهميتها بالنسبة له، واختيار السمات الشخصية والاجتماعية التى يرى أنها ضرورية. وتستند بعض هذه الاستبيانات وموازين التقدير إلى أن هناك أنماط ميول متماثلة تُميز الأفراد الناجحين فى مهنة معينة أو عمل محدد كالتدريس مثلاً، كما أن هناك أنماط ميول فارقة بين الأفراد الناجحين فى مهن مختلفة كالهندسة والطب. وبالطبع تقيس هذه الاستبيانات عينة محددة من ميول الفرد، حيث إن كثيرا من الاستبيانات المتوافرة فى الوقت الحاضر تفيد فى التوجيه المهنى، بينما عدد قليل منها يمكن الإفادة منه فى التوجيه التربوى.

أساليب بناء استبيانات الميول :

تختلف أساليب بناء استبيانات الميول Interest Inventories وتقدير درجاتها Keying عن الأساليب المتعارف عليها في قياس الجوانب المعرفية، وذلك لأن مفردات الاختبارات المتعلقة بهذه الجوانب يكون لها إجابة صحيحة واحدة مقبولة، بينما مفردات مقاييس الجوانب غير المعرفية أو الوجدانية بعامة ليس لها إجابة صحيحة محددة.

إذ ربما تُعد استجابة الفرد لإحدى فقرات استبيان للميول صحيحة بالنسبة لأعضاء مهنة معينة بينما تُعد الاستجابة خطأ بالنسبة لأعضاء مهن أخرى. ويعتمد تفسير درجات هذه الاستبيانات على الأسلوب الذى يُتبع فى بنائها.

ويوضح الشكل التخطيطى (١١ ـ ٢) التالى أهم هذه الأساليب :



شكل (١١ ـ ٢) يوضح أهم أساليب بناء استبيانات الميول

وسوف نوضح فيما يلى كلا من هذه الأساليب بمزيد من التفصيل:

أسلوب يستند إلى أسس نظرية أو منطقية : استُخدم هذا الأسلوب فى مستهل نشأة بناء استبيانات الميول والشخصية، حيث تُنتقى فقرات أو مفردات الاستبيان استنادًا إلى التحليل النظرى أو المنطقى للسمة أو الخاصة المراد قياسها، وهذا يماثل الأسلوب الذى يُتبع فى تقدير صدق محتوى الاختبارات الذى أوضحناه فى الفصل الخامس. وهذا يتطلب من القائم ببناء استبيان الميول أن يحدد الأساس المنطقى الذى يُستند إليه فى تصنيف الميول وبناء الفقرات التى تبدو أنها تقيس المجالات المختلفة.

وقد استخدم لى وثروب Lee - Thrope هذا الأسلوب فى بناء استبيان الميول المهنية الذى سوف نوضحه فيما بعد. ويتم تقدير درجات الفقرات على أساس منطقى. وعلى الرغم من أنه ربما لا يكون هناك اتفاق عام على تقدير درجة فقرة ما، إلا أنه عند

بناء المفردات ينبغي الاهتمام بانتقاء الفقرات التي يسهل تقدير درجاتها. فمثلاً تقدير درجة الفقرة «أميل إلى قراءة المقالات الاجتماعية» ربما يكون أيسر من تقدير درجة الفقرة «أفضل المشاركة في الحفلات الاجتماعية مع الأصدقاء» عن المشاركة في الحفلات الجماعية الكبيرة». وعمومًا فإن بناء الاستبيان استنادًا إلى الأسلوب المنطقي يُعد خطوة أولية يمكن أن يتبعها خطوات تحليلية إحصائية أو إمبيريقية أخرى لتنقية الفقرات، وذلك لأن الفقرات التي تبدو أنها تقيس سمة أو مجالا معينًا ربما لا تقيسه بالفعل. لذلك لا يوصى بالاقتصار على هذا الأسلوب في بناء الاستبيانات.

أسلوب يستند إلى بناء فقرات متجانسة: يُعد هذا الأسلوب Homogeneous Keying امتدادًا واستكمالاً للأسلوب السابق. فبعد انتقاء فقرات الاستبيان على أساس منطقي وتصنيف هذه الفقرات في مجموعات متجانسة تقيس مجالات مختلفة مثل الميل الخلوي، والميل إلى مهنة معينة، والميل إلى الأعمال الإدارية مثلاً، يمكن باستخدام الأساليب الإحصائية التي أوضحناها في الفصل الرابع التحقق من الاتساق الداخلي أو تجانس الفقرات.

إذ يرى كثير من علماء القياس أن المقاييس المتسقة داخليًا أو متجانسة هي التي تُضفي معنى على المقياس وتُمكننا من تفسير درجاته. فالفقرات التي يكون ارتباطها بغيرها من الفقرات التي يشتمل عليها المقياس منخفضًا بدرجة واضحة يجب حذفها لأنها لا تنتمي إلى مجموعة الفقرات التي تقيس سمة أو مفهومًا معينًا. وهذا يماثل أسلوب التحقق من صدق التكوين الفرضي الذي أوضحناه في الفصل الخامس.

ويمكن أيضًا إخضاع مصفوفة الارتباطات بين الفقرات المختلفة التي يشتمل عليها الاستبيان للتحليل العاملي Factor Analysis للتوصل إلى أقل عدد من العوامل التي تفسر هذه الارتباطات. وبذلك نحصل على تجمعات من الفقرات التي يقيس كل منها عاملاً معينا يمكن تسميته بحسب محتوى فقرات كل تجمع، وهذه العوامل أو السمات يمكن قياسها. ففي مجال الميول ربما ترتبط مجموعة من الفقرات فيما بينها ارتباطًا مرتفعًا، وبذلك تقيس عاملاً أحادي البعد يمكن أن يطلق عليه مثلاً الميل الميكانيكي.

وينطبق ذلك أيضا على مقاييس الاتجاهات واستبيانات الشخصية كما سيتضح في الفصلين التاليين. وقد استخدم جيلفورد ومعاونوه .Guilford et al هذا الأسلوب في بناء استبيان مسح الميول، كما استخدمه كيودر Kuder في بناء استبيان المسح العام للميول.

أسلوب يستند إلى التحقق الإمبيريقي، يعتمد هذا الأسلوب في انتقاء فقرات استبيانات الميول على درجة تمييزها بين مجموعة من الأفراد تنتمي إلى مهنة

معينة ومجموعة من عامة الناس أو مجموعات من مهن مختلفة. فالمجموعة الأولى تكون استجاباتها لفقرات الاستبيان مختلفة عن استجابات عامة الناس، أو أفراد المهن الأخرى للفقرات نفسها. فعندئذ يمكن تضمين هذه الفقرات في استبيان الميول، ويتم تقدير درجات فارقة تتناسب مع اختلاف استجابات هذه المجموعات.

ويتطلب هذا الأسلوب أيضًا تجميع عدد كبير من الفقرات وتجريبها ميدانيًا على مجموعة من الأفراد الناجحين في مهنة معينة مثلاً ومجموعة أخرى من عامة الناس. ويمكن تعريف الفرد الناجح بأنه الذى قضى في مجال عمله عددًا معينًا من السنوات على الأقل، ويواظب على هذا العمل، ويفضله عن غيره من الأعمال، ويحصل على تقديرات متميزة من مشرفيه، ولا يقل عمره أو يزيد عن عمر معين. أى أنه ينبغى تعريف محك النجاح الذى سيتم التمييز في ضوئه نظرًا لأن هذا الأسلوب لا يستند إلى افتراضات معينة تتعلق بسمة أو خاصة من خصائص أفراد المجموعات المختلفة، وإنما يتم انتقاء الفقرات التى تميز بين هذه المجموعات، فالانتماء إلى مجموعة معينة هو المحك الذى يقدَّر في ضوئه صدق الاستبيان. ويستخدم هذا الأسلوب أيضًا في بعض استبيانات الشخصية كما سيتضح في الفصل الثالث عشر.

ولعل هذا الأسلوب يتميز بأنه لا يُمكِّن المستجيبين من تزوير استجاباتهم لأن استجابات المجموعة أو المجموعات التى يُستند إليها كمحك لا تكون معلومة لهم. غير أن اختيار الفرد الناجح في مهنة معينة يعتمد على تعريف معين ربما لا يكون كافيًا أو شاملاً، فعندئذ لا يكون الاستبيان مناسبًا، كما أن هذا الأسلوب في بناء الاستبيانات ربما يحقق الصدق التلازمى. ولكن يبرز تساؤلان حول إمكانية الإفادة من الاستبيان في التنبؤ بمحك مستقبلى وعدم الاقتصار على تقييم انتماء الفرد لمجموعة مهنية معينة، وكذلك مدى تعميم نتائج الاستبيان، أى التوصل إلى فهم متكامل للسمة المقاسة في ضوء إطار نظرى معين استنادًا إلى نتائجه. فالتساؤل الأول يتعلق بالصدق التنبؤى الذى يمكن التحقق منه إمبيريقيا، أما التساؤل الثانى فيتعلق بصدق التكوين الفرضى الذى يتطلب أدلة متعددة.

لذلك يتضح أنه على الرغم من تعدد أساليب بناء استبيانات الميول والشخصية، إلا أن هناك تكاملا فيما بينها، فالأسلوب المنطقى **Rational Keying** يُعد خطوة أولية لانتقاء مجموعة من الفقرات. ويتم تجريب هذه الفقرات على جماعة مرجعية مناسبة ممثلة للمجتمع الذى سيطبق عليه الاستبيان فيما بعد، وتكوين مصفوفة الارتباطات بين درجات الفقرات، وإجراء التحليل العاملى من أجل التوصل إلى مجموعات متجانسة

من الفقرات وبعد ذلك يمكن التحقق من الصدق المرتبط بمحك بالأساليب الإمبريقيةEmperical Keying، وبذلك يكون للاستبيان أساس نظرى ومنطقى، وفى الوقت نفسه تتأكد فائدته العملية نتيجة استخدام الأسلوب الإمبيريقى.

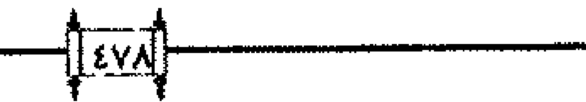
بعض المشكلات المتعلقة بقياس الجوانب غير المعرفية :

إن قياس الجوانب غير المعرفية أو الوجدانية للسلوك يُعد أكثر صعوبة من قياس الجوانب المعرفية لأسباب كثيرة نذكر منها:

(١) عدم الاتفاق التام على تعريف مكونات الجوانب الوجدانية للسلوك، كما أن هذه المكونات تصعب ملاحظتها مباشرة فى سلوك الأفراد، إذ كثيراً ما يجد المعلم أو المرشد التربوى أو المهنى صعوبة فى تحديد ما يود ملاحظته فى سلوك الفرد الذى لديه ميل نحو نشاط أو عمل ما، أو اتجاه معين. فمما لا شك فيه أن هذه المكونات الوجدانية تحدث داخل الفرد. وقياس وتقويم هذه المكونات غير المنظورة يحتاج إلى تحليل سلوكى بحيث يمكن ملاحظة السلوك ملاحظة مباشرة تعكس تأثير هذه المكونات. وتتباين وجهات النظر تبايناً ملحوظاً حول إجراءات ونواتج هذا التحليل المتعلق بالأداء المميز للفرد.

(٢) يصعب قياس بعض الجوانب الوجدانية ، حيث إنها ربما تتطلب ملاحظة الفرد ملاحظة متكررة فى مواقف مختلفة يمكن أن تظهر فيها الاستجابات التى نهتم بملاحظتها، نظراً لأن ملاحظته فى موقف واحد لا يكفى للحكم على أدائه المميز. فالسلوك الوجدانى يعتمد على الحالة المزاجية للفرد، وألفته بالموقف، وظروف البيئة المحيطة به، وغير ذلك من المتغيرات. فالفرد الذى يبدو عدوانيًا في بعض المواقف ربما يجعلنا نحكم بأنه شخص عدوانى، ولكن ربما يكون فى حالة من الضيق نتيجة قلقه المستمر على شيء ما، ويبدو فى حالة تكيف تام بعد انقضاء عدة أسابيع أو عدة شهور، وبذلك فإن ما نلاحظه فى مواقف قليلة عن فرد معين ربما يكون عرضًا وقتيًا يميز الفرد فى هذا الوقت.

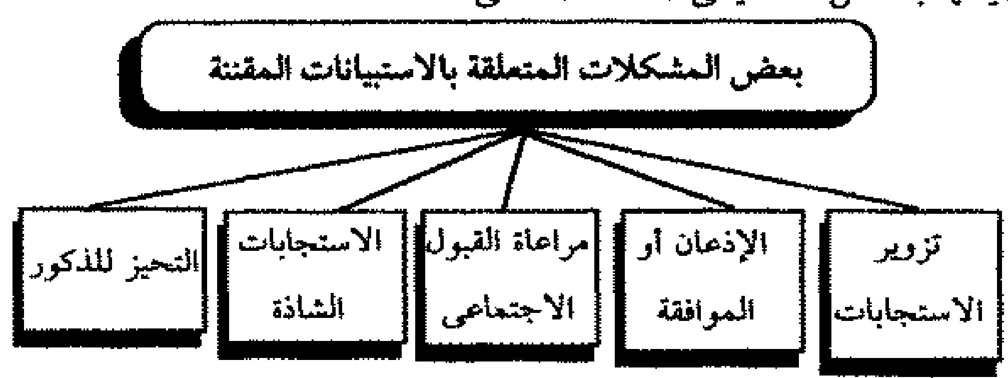
(٣) نظراً لصعوبة التحديد الفعلى الدقيق لكثير من المفاهيم والمكونات الوجدانية، فإننا لا نستطيع الاعتماد على أداة قياس واحدة. إذ من الصعب احتواء هذه المكونات جميعًا وقياسها مباشرة بأداء قياس واحدة؛ لذلك ربما يفضل اتباع استراتيجية الالتفاف حول هذه المكونات، وذلك باستخدام مقاييس متعددة مناسبة، وبذلك يمكن أن نحصل على صورة أفضل عن المكونة المراد قياسها.



(٤) إن قياس الجوانب غير المعرفية أو الوجدانية يشمل الفروق بين الأفراد وكذلك التغيرات التى تحدث فى السلوك عبر الزمن، وهذا يُلقى عبئًا كبيراً على القائم ببناء مقاييس هذه الجوانب فى التحقق من صدق وثبات البيانات المستمدة من هذه المقاييس.

بعض المشكلات المتعلقة بالإستبيانات المقننة للميول :

سوف نناقش فيما يلى بعض أهم المشكلات المتعلقة بالاستبيانات المقننة لقياس الميول، وتنطبق هذه المشكلات أيضًا على استبيانات الاتجاهات والشخصية، ويمكن توضيحها بالشكل التخطيطى (١١ ـ ٣) التالى :



شكل (١١ ـ ٣) يوضح بعض المشكلات المتعلقة بالاستبيانات المقننة

تزييف أو تحيز الاستجابات : تُعد مشكلة تزييف أو تحيز الاستجابات Faking فى استبيانات الميول وغيرها من استبيانات الجوانب غير المعرفية من المشكلات التى اهتمت كثير من البحوث بدراستها ، وبخاصة بعد دراسة كرونباك Cronbach التى قدم فيها مراجعة شاملة لهذه المشكلة (Cronbach , 1950). ففى اختبارات الذكاء والاستعدادات الخاصة لا يلجأ الفرد إلى تزييف استجاباته، وإنما يمكن أن يُخمِّن الإجابة الصحيحة، أو يختار إجابات خطأ فى مفردات الاختيار من متعدد، أو يترك المفردة دون إجابة. ولكن فى استبيانات الجوانب غير المعرفية يستطيع الفرد تزييف إجابته بطرق متعددة.

لذلك اهتم الباحثون بدراسة هذه المشكلة المعقدة. فعلى الرغم من أن الاستبيانات التى تقيس هذه الجوانب تهدف لمعاونة الفرد فى تعرُّف ميوله واتجاهاته وسمات شخصيته وتوجيهه تربويًا ومهنيًا استنادًا إلى نتائجها، إلا أن بعض الأفراد يلجأون إلى تزييف استجاباتهم لصالحهم، وبذلك لا تُعبر هذه الاستجابات عن أدائهم المميز الفعلى. ولعل هذا يعتمد على الغرض من استخدام استبيان الميول أو

الشخصية، فإذا كــان الغرض انتقاء الأفراد لوظيفة أو مهنة شاغــرة، فإن احتمال تزييف فرد معين لاستجاباته لكى يعطى انطباعا مقبولا عن نفسه يكون أكبر مما لو كان الغرض من الاستـبيان التوجيـه المهنى. بل ربمــا أيضًا فى التوجـيه التربوى يلجـأ الطالب إلى الكذب فى إجاباته على استبيان الميــول لكى يبين أنه يميل إلى دراسة الطب أو الهندسة مثلاً.

وعلى العكس من ذلك، فإن هناك بعض الأفراد يتعـمّدون إظهار أنفـسهم بمظهر عدم التكيف أو أنهم غير أسوياء للتهرب من الالتحاق بالجيش أو الحجز بـمستشفيات الأمراض العقلية، أو مـرتكبى الجرائم، ولكن ليس معنى ذلك أن تـزييف الاستجابات يكون مُتعمدًا فــى جميع الأحوال، فبعض علماء النفس يــرون أن الفرد ربما يكون لديه دوافع أو نزعــات شبــة شعـورية Subconscious للاسـتـجابة بطريـقة أو نمط مـعين. فالمراهـق الذى يرفض سلطة الوالدين ويــودُّ والداه إلحاقه بكليـة معينة ربمــا يلجأ إلى الكذب فى استبيان المـيول وبخاصة فى المقياس المتعلق بالميــل نحو المهنة التى تُعد هذه الكلـية لها. لذلــك ينبغى تـعرُّف دوافع الفرد ونـزعاته عند تفـسير الـدرجات فى استبيانات الميول وغيرها من الاستبيانات.

الاتجاه العقلى فى الاستجابة : إن الاستبيانات أو التقارير الذاتية ليست عُرضة فقط لتزييف الاستجابات أو تحيزها، وإنما أيضًــا الدرجات المستمدة منها عرضة للتأثر باتجاهات عقلية مـعينة فى استجابات الفرد، وبخاصة إذا كانت تـشتمل على فقرات أو عبارات غـامضة. وهذه الاتجاهات الـعقلية ربما تـرجع إلى نزعات شبه شـعورية لدى الفرد للاسـتجـابة بطريـقة أو توجُّه معين بغض النظر عن مـحتوى فقـرات أو عبارات الاستبيان. ويُطلق على هذه النزعات الاتجـاه العقلى الـمعين فى الاستجابة Response Sets. وهذا الاتجاه العقلى يجعل الفرد يستجيب لفقرات الاستبيان بطريقة تختلف عن استجابتــه لو قدّمت له الفقرات بشكل آخر أو فــى مضمون مختلف. فالاســتبيان الذى يُفترض أنه يقيس سمة أو خاصــة معينة لدى الـفرد مثل الميول ربما يقيس أيضًا خاصة أخرى أى الاتجاه العقلى، وهذه الخاصة الأخرى ربما لا ينعكس تأثيرها فى درجته لو أننا أحدثنا تغييرًا فى الموقف الاختبارى بعامة، مثل نمط أو شكل فـقرات الاستبيان، وأسلوب صيــاغتها، وعــدد وطبيعة الـبدائل التى تشـتمل عليهــا الفقرات، وتعلـيمات الاستبيان، والغرض الذى سوف تُستخدم فيه درجات الاختبار، أو القرارات التى سوف تستند إلى هذه الدرجات.

ويوجد نمطان رئيسان لتباين الاتجاه العقلي: أحدهما انتقالي ويقتصر على استبيان معين أو موقف اختباري معين، والآخر مستقر ومتسق ويتصف بالعمومية إلى حد ما، أي يؤثر في استجابات الفرد في الاستبيانات والمواقف الاختبارية المختلفة. ويوجد أيضًا نمطان لهذا الاتجاه العقلي المستقر: أحدهما يعكس فروقًا فردية ثابتة نسبيًا ولكنها غير ذات أهمية مثل استخدام اللغة أو العادات التي تبدو من تعبيرات الفرد، والآخر يعكس جوانب مهمة للشخصية.

لذلك فإن التباين الذي يُعزى إلى هذا النمط الثاني ربما يُعد مؤشرًا لخصائص الشخصية؛ ولذلك يطلق عليه أسلوب أو نسق الاستجابة Response Style. وهذا الأسلوب أو النسق يعني اتساق طريقة الاستجابة لشكل ما من أشكال فقرات الاستبيانات بغض النظر عن محتوى نوعي معين.

ويتميز هذا الأسلوب أو النسق بالاستقرار النسبي عبر الزمن ويُسهم في تعميم الاستجابات سواء في المواقف الاختبارية أو غير الاختبارية، وهذا الأسلوب ينبغي دراسته في حد ذاته (Jackson & Messick , 1958).

وسوف نعرض فيما يلي الأنواع الرئيسة للاتجاهات العقلية Response Sets التي أشارت إليها معظم أدبيات القياس التربوي والنفسي:

الإذعان أو الموافقة : اهتمت كثير من الدراسات بهذا النوع من تحيُّز الاستجابات أو الاتجاه العقلي. ويُقصد بالإذعان أو الموافقة Acquiescence نزعة الفرد في استجابته على فقرات استبيانات الميول أو الشخصية إلى الموافقة بغض النظر عما تشير إليه الفقرات. أي أن الفرد ينزع إلى الاستجابة دومًا إما بـ «نعم» أو «لا»، أو «موافق» أو «غير موافق». فإذا ما طبق استبيان للميول أو الشخصية على هؤلاء الأفراد، فإن الدرجات التي يحصلون عليها لا تُعبر عن السمة المقاسة لديهم بقدر ما تعكس هذا الاتجاه العقلي، وبذلك لا نستطيع الاستناد إلى هذه الدرجات في اتخاذ قرارات مهمة بشأنهم.

وهناك أدلة بحثية متعددة أشارت إلى أن الإذعان أو الموافقة تُفسِّر جزءًا كبيرًا من تباين استبيانات الجوانب غير المعرفية، وبخاصة استبيانات الشخصية. وقد سبق أن أوضحنا أن بعض علماء النفس يرون أن هذا الاتجاه العقلي ربما يُعد سمة شخصية عند

الأفراد الذين لديهم نزعة شبه شعورية للاستجابة في معظم الأحيان لفقرات استبيانات الميول والشخصية بالموافقة، فهم يرون أن مثل هؤلاء الأفراد يكونون منقادين، وممتثلين، دائمًا وينسحبون من التحديات في حياتهم اليومية.

ويقترح سميث (Smith , 1961) أنه يمكن التحكم جزئيًا في هذا الاتجاه العقلي في الاستجابة بالعناية بصياغة فقرات الاستبيان، فالفقرات التي تشتمل على عبارات متطرفة ربما لا تؤدى إلى استجابات الموافقة أو الإذعان بنفس قدر العبارات المعتدلة. كما يرى إيليوت (Elliott , 1961) أن هناك ثلاثة محددات للإذعان أو الموافقة هي: نوع أو شكل الفقرة (عبارات موجبة، وعبارات سالبة، واختيار من متعدد) ، ومدى تَعلُّق محتوى الفقرة بسمة شخصية، والمستوى العقلي للفرد المستجيب، ويرى أن هذه المتغيرات الثلاثة متفاعلة.

مراعاة القبول الاجتماعي : يُعد هذا الاتجاه العقلي في الاستجابة Social Desirability نوعًا من التحيز، حيث يحاول الفرد إظهار نفسه بصورة مقبولة اجتماعيًا في استجابته لفقرات معينة في استبيانات الميول والشخصية، أي أن الفرد يحاول أن يُظْهِر نفسه مثلاً كإنسان متكيف ومنفتح ومتفاعل مع الآخرين. فكثير من الأفراد يوافقون على عبارة مثل «أحاول قدر استطاعتي تجنب إثارة المشاكل»، أو «أفضل الأعمال التي تقدِّم خدمات إنسانية»، أو «أميل إلى مساعدة الآخرين». وهذه الموافقة تتم بطريقة شبه شعورية، حيث إن الفرد يحاول أن يكون مقبولاً اجتماعيًا، وبذلك لا تُعد استجابته نوعًا من الكذب أو الخداع لأن الفرد لا يتعمَّد ذلك.

ويعتمد هذا الاتجاه العقلي على محتوى فقرات مقاييس التقدير الذاتي مثل الاستبيانات. فالفرد لا يستطيع تزييف استجابته الحقيقية لكي تكون مقبولة اجتماعيًا ما لم يدرك العلاقة بين محتوى الفقرة التي يستجيب لها والطريقة المحتملة لتقدير درجتها.

ولعل تبرير هذا الاتجاه العقلي أكثر وضوحًا من الاتجاهات العقلية الأخرى. فهنا ربما يكون التبرير هو تجنب الفرد للنقد أو التهديد الخارجي أو أن يتضح سوء تكيفه.

وقد أوضحت دراسة كراون ، ومارلو (Crown & Marlowe , 1964) ودراسة فريدريكسن (Fredrikson , 1965) أن شدة الاتجاه العقلى المتعلق بمراعاة القبول الاجتماعى يرتبط بحاجة الفرد إلى حماية نفسه، والامتثال لما هو سائد فى المجتمع.

ولقد كان إدواردز (Edwards , 1957) أول من اهتم ببحث متغير «مراعاة القبول الاجتماعى»، وبيّن أن هذا المتغير أحادى البُعد. فإذا طلبنا من عدد من المحكمين تقدير درجة القبول الاجتماعى لكل عبارة من العبارات التى يشتمل عليها استبيان معين، فإنه يمكن الحصول على ميزان أو مقياس متدرج أحادى البُعد لهذا الاتجاه العقلى.

كما أجرى إدواردز (Edwards , 1957) ، وكراون ومارلو (Crown & Marlowe, 1960 ,1964) بحوثًا متعددة حول تأثير «مراعاة القبول الاجتماعى» فى درجات استبيانات التقرير الذاتى، وأعدّوا موازين يمكن استخدامها فى تقييم النزعة العامة للفرد لإظهار نفسه بصورة مقبولة اجتماعيًا. كما يمكن استخدام هذه الموازين أثناء بناء استبيانات الجوانب الوجدانية للتحكم فى هذا الاتجاه العقلى فى استجابات الأفراد، وذلك بحذف الفقرات التى ترتبط بمقياس مراعاة القبول الاجتماعى.

ومن أجل تقليل تأثير هذا الاتجاه العقلى فى الاستجابات اقترح إدواردز Edwards استخدام استبيانات تتطلب الاختيار الجبرى Forced-Choice Inventories ففى هذه الاستبيانات تُقدم أزواج من العبارات المقبولة اجتماعيًا بدرجة متكافئة ومتباينة فى محتواها وصدقها، أى أن كلا من العبارتين ترتبط ارتباطًا متمايزًا بمحكات إمبيريقية مناسبة. فإذا اختار الفرد إحدى العبارتين التى يرى أنها تنطبق عليه بدرجة أفضل يحصل على (درجة واحدة) إذا ارتبطت العبارة بالمحك، بينما يحصل على (صفر) إذا اختار عبارة لا ترتبط بالمحك.

ويمكن أن يُقدَّم أى عدد من بدائل العبارات لكل فقرة من فقرات الاستبيان بدلا من زوج واحد. وسوف يتضح ذلك فى عرضنا لأمثلة من استبيانات الميول، وبخاصة استبيان كيودر حيث استخدم الفقرات جبرية الاختيار. وعلى الرغم من أن هذا النوع من الفقرات يزودنا بمعلومات أكثر تفصيلاً من الفقرات المألوفة، إلا أنه يمكن تزييف استجاباتها أيضًا. إذ من الصعب الحصول على زوج من العبارات متماثل تمامًا، والفرد الذى يود إعطاء صورة أفضل عن نفسه يمكن أن يتعرف أحد البديلين الذى يحتمل بدرجة أكبر أن يكون مقبولاً.

كما أن تطبيق هذا النوع من الفقرات يتطلب وقتًا أطول للحصول على عدد من الاستجابات مساويًا للعدد الذى نحصل عليه إذا استخدمت فقرات غير جبرية الاختيار.

الاستجابات الشاذة : يميل بعض الأفراد إلى الاستجابة الشاذة أو غير الشائعة Deviant (Atypical) Responses على فقرات استبيانات الجوانب الوجدانية، وبخاصة مقاييس الشخصية، وهذه النزعة تكون متسقة عبر فقرات الاستبيان وبمرور الزمن، فالأشخاص غير الأسوياء يستجيبون لهذه الفقرات استجابة شاذة بغض النظر عن محتواها.

ومعظم الدراسات التى اهتمت بدراسة هذا الاتجاه العقلى طلبت من هؤلاء الأفراد أن يختاروا استجاباتهم لفقرات استبيان معين على ميزان تقدير يتراوح بين: أهتم جدًا أو أوافق جدًا، ولا أهتم ، أو لا أوافق على الإطلاق. وهنا يلجأ هؤلاء الأفراد إلى اختيار استجابات تختلف عن الاستجابات الشائعة لجماعة مرجعية معينة. وقد قدَّم سيشرست وجاكسون (Sechrest & Jackson, 1963) تحليلاً تفصيليًا لنزعات الاستجابات الشاذة وكيفية قياسها وتفسيرها.

التحيز للذكور : وُجِّه كثير من النقد لاستبيانات الميول فى مستهل نشأتها بحجة أنها تتحيز للذكور، وتُسهم فى توجيه الإناث إلى أعمال ومهن تتفق مع الأدوار التقليدية للمرأة، مثل أعمال السكرتارية، والتمريض، والتدريس للأطفال، والحرف اليدوية البسيطة. لذلك فإن استبيان سترونج للميول المهنية فى طبعته الأولى كانت له معايير منفصلة لكل من الذكور والإناث. وقد طالبت العديد من المؤسسات الأمريكية التى تدافع عن حقوق المرأة بإعادة النظر فى هذه الاستبيانات بحيث تكشف عن ميول المرأة نحو المهن العليا التى ترفع من مستواها الاجتماعى والاقتصادى.

لذلك أُعيد النظر فى استبيانات الميول بحيث لا يكون هناك استبيان خاص بالذكور وآخر خاص بالإناث، وإنما يُستخدم مقياس واحد ومعايير واحدة لكل منهما. غير أن هذا التطوير قد أسهم فى تقليل صدق استبيانات الميول، نظرًا لأن الواقع الاجتماعى والنفسى يشير إلى اختلاف ميول الذكور عن الإناث (Campbell , 1977).

لذلك يُلاحظ أن بعض الاستبيانات عادت مرة أخرى إلى التمييز بين معايير الذكور والإناث فى المقاييس المتعلقة بالميول المهنية، ولكنها دمجت معايير كل منهما فى مقاييس الميول العامة، وعمومًا فإن هذا ربما يُقلل من التحيز لأحد الجنسين ولكنه لا يستبعده تمامًا.

تعقيب على مشكلة تحيز الاستجابات :

لاحظنا تعدد مصادر الاتجاهات العقلية Response Sets التى يمكن أن تسهم فى تحيز استجابات الأفراد فى مقاييس التقرير الذاتى مثل الاستبيانات. غير أنه لا يوجد



اتفاق حول مدى استقلالية كل من هذه المصادر عن الآخر. وعلى الرغم من أنه يمكن التنبؤ بالاتجاه العقلى المتعلق بالإذعان أو الموافقة من معرفتنا بالاتجاه العقلى المتعلق بمراعاة القبول الاجتماعى، مما يعنى أن المصدرين يعبران عن ظاهرة واحدة، إلا أنه لا يوجد اتفاق أيضًا بهذا الشأن.

فالدراسة التى أجراها كاوش وكينيستون (Couch & Keniston 1960) حول «الإذعان أو الموافقة» كمتغير من متغيرات الشخصية انتقدها كل من إدواردز ، ووالكر (Edwards & Walker , 1961) ، وتايلر (Tyler , 1961) على أساس أن نتائجها يمكن تفسيرها فى ضوء «مراعاة القبول الاجتماعى» بدلاً من «الإذعان»، ولكن عندما أجرى الباحثان الأصليان تحليلاً عامليًا لبيانات دراستهما توصلاً إلى أن مراعاة القبول الاجتماعى، والإذعان يُعدان عاملين مختلفين، بينما توصل إدواردز، ووالكر من نتائج تحليل عاملى آخر أن المصدرين يجمعهما عامل مشترك.

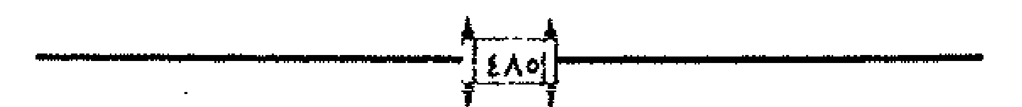
وإذا سلَّمنا بأن «الإذعان أو الموافقة» يُعد مصدرًا مستقلاً، فإن هناك أيضًا عدم اتفاق حول أهميته النسبية مقارنة بمصدر «مراعاة القبول الاجتماعى»لذلك فإن الدراسات لم تتفق على أى المصادر تُسهم فى التباين الكلى لدرجات الاستبيانات إسهامًا أكبر. ولكن أوضح بعضها أن مصادر الاتجاهات العقلية فى الاستجابات تُسهم فى التباين الكلى للدرجات أكثر مما تُسهم به السمات التى يُفترض قياسها بهذه الاستبيانات، (Keniston & Messick , 1961).

غير أن رورار (Rorrer , 1965) قام بمراجعة شاملة للدراسات فى هذا المجال واستنتج أن أساليب أو أنساق الاستجابات Response Styles نادرًا ما تُفسر أكثر من ٢٠٪ من تباين درجات الاستبيانات، مما يجعل اعتقادنا بأن الأفراد يعتمدون فى استجاباتهم لاستبيانات الميول والاتجاهات والشخصية على هذه الأساليب أو الأنساق غير صحيح.

وعلى أى حال فإنه ينبغى أن نحتاط للآثار المحتملة لهذه النزعات فى بناء وتفسير نتائج هذه الاستبيانات فى ضوء المقترحات التالية:

(أ) تكوين علاقات وديّة مع المختبرين لإقناعهم بالفائدة التى سوف تعود عليهم مستقبلاً نتيجة تحرّى الصدق فى استجاباتهم وتشجيعهم على ذلك.

(ب) محاولة إخفاء الهدف من الاستبيان عند بنائه بحيث يبدو أن محتوى الاستبيان يقيس شيئًا مختلفًا عما يراد قياسه. غير أن هذا الاقتراح يناسب بعض الاستبيانات دون الآخر، كما أنه يؤدى إلى مشكلات عملية وأخلاقية،



وربما يؤدى كذلك إلى عدم ثقة المستجيب فى المرشد التربوى أو النفسى مما يسيء إلى علاقة المودة بينهما .

(جـ) صياغة الفقرات صياغة واضحة ومحايدة، وإعداد مفاتيح تصحيح خاصة لكشف هذه الاتجاهات العقلية فى الاستجابات، كما فى مقياس كيودر للميول وكما سيتضح فى الفصل الثالث عشر، والعناية بإعداد تعليمات الاستبيانات.

(د) استخدام فقرات من نوع الاختيار الجبرى كما أشرنا منذ قليل، غير أن هذا لا يكشف عن الاتجاهات العقلية وإنما يقلل من أثرها إلى حد ما .

مشكلات تفسير درجات استبيانات الميول:

تعتمد الاختبارات التى تقيس الجوانب المعرفية فى تفسير درجاتها على المعايير أو محكات الأداء. وعلى الرغم من أن مقاييس الجوانب غير المعرفية تعتمد أيضًا على الإطار المرجعى النسبى (المعايير) ، والإطار المرجعى المطلق (محكات أو مستويات الأداء) فى تفسير درجاتها، إلا أن بعض هذه المقاييس، وبخاصة استبيانات الميول والشخصية، تعتمد فى ذلك على الإطار المرجعى الفردى Ipsative Scales .

فالدرجة فى ضوء هذا الإطار المرجعى تصف مستوى الفرد فى أحد المتغيرات أو السمات بالنسبة لمستواه فى متغير آخر أو سمة أخرى، أى أن تفسير الدرجات فى هذه الحالة يركز على مقارنة الفرد بنفسه فى أكثر من متغير. وبالطبع يختلف ذلك عن تفسير الدرجات فى ضوء الإطار المرجعى النسبى الذى يقارن درجة الفرد بدرجات أقرانه فى متغير أو أكثر، أو تفسير الدرجات فى ضوء الإطار المرجعى المطلق الذى يقارن درجة الفرد بمحك أو مستوى أداء محدد مسبقًا، كما سبق أن أوضحنا فى الفصول السابقة.

ففى ضوء الإطار المرجعى الفردى يمكننا تفسير درجات الفرد فى استبيان للميول بأن نقول أن درجة فرد فى الميل العلمى ٧٠ ، ودرجته فى الميل الكتابى ٣٠، لذلك فإن ميله العلمى أقوى. أما التفسير فى ضوء الإطار المرجعى النسبى فيكون مختلفًا ، فإذا حصل فرد على الدرجة ٧٠ فى اختبار ما وحصل فرد آخر على الدرجة ٣٠ فى الاختبار نفسه ، فإن هذا يعنى أن الفرد الأول أفضل من الفرد الثانى فيما يقيسه الاختبار.

وتتميز درجات المقاييس التى تُنسب إلى هذا الإطار المرجعى الفردى Ipsative Scales بأن قيم متوسطات الدرجات فى جميع المقاييس تكون متساوية لكل فرد، فمثلاً يمكن أن تكون درجة فرد ما فى مقياسى الميل العلمى والكتابى ٧٠ ، ٣٠ بينما

درجة فرد آخر في المقياسين ٩٠ ، ١٠ على الترتيب، حيث يظل المتوسط ٥٠ لكل منهما. ويمكن الحصول على الدرجات التي تفي بهذه الخاصة من الاختبارات والمقاييس التي يتبع في بنائها أسلوب معين يؤدي مباشرة إلى مثل هذه الدرجات فردية المرجع.

كما يمكن الحصول عليها باستخدام طرق إحصائية لتحويل الدرجات المعيارية أو المئينيات وغيرها من الدرجات المُحولة إلى درجات تُنسب إلى الإطار المرجعي الفردي. فالدرجات الخام في المقاييس الفرعية لاستبيانات الميول تُحول أولاً إلى درجات معيارية ، وبذلك يتساوى متوسطها وانحرافها المعياري، وبعد ذلك يُضاف مقدار ثابت إلى درجات كل فرد لكي يصبح مجموع درجات كل منهم متساويًا. فمثلاً إذا كان لدينا أربعة متغيرات، وكان المجموع الكلي لدرجات فردين في هذه المتغيرات ٢٠٠ ، ٢٨٠ ، فإنه ربما نُضيف مقدارًا ثابتًا لكل درجة من درجات الفرد الأول في هذه المتغيرات وليكن ٢٥ ، ونُضيف مقدارًا ثابتًا آخر لكل درجة من درجات الفرد الثاني في هذه المتغيرات أيضًا وليكن ٥ بحيث يصبح مجموع الدرجات الكلية لكل منهما في هذه المتغيرات ٣٠٠ .

وبذلك تتم المقارنة داخل الأفراد، فكل فرد ربما يحصل على درجات مرتفعة في بعض المتغيرات، ودرجات منخفضة في المتغيرات الأخرى، أو ربما يحصل على درجات مساوية لقيمة المتوسط في جميع المتغيرات. أي أن الفرد لا يمكن أن يحصل على درجات مرتفعة أو درجات منخفضة في جميع المتغيرات.

وهذا التفسير الذي يُنسب إلى الفرد ذاته يمكن الإفادة منه في مواقف متعددة، بل أحيانًا يكون ضروريًا، ففي مواقف الإرشاد والتوجيه وما شابه ذلك، حيث يتركز الاهتمام على الفرد وتحديد ميوله الفعلية والمقارنة بين جوانب قوته وضعفه في التحصيل والقدرات، تُعد الدرجات التي تنسب إلى الإطار المرجعي الفردي Ipsative Scores أبسط الطرق التي تُزودنا بهذه المعلومات مباشرة. كما أن هذه الدرجات تفيد في أساليب التقدير الذاتي التي تستند إلى استبيانات الاختيار الجبري Forced - Choice Items من أجل تقليل أثر الاتجاه العقلي المتعلق «بمراعاة القبول الاجتماعي» وغيره من الاتجاهات العقلية في الاستجابات، وبذلك تُسهم في زيادة صدق الاستبيان بدرجة أكبر مما لو استندت الدرجات إلى معايير جماعة مرجعية معينة.

غير أنه ينبغي ملاحظة أن تفسير درجات استبيانات الميول استنادًا إلى الإطار المرجعي الفردي لا يعكس مدى عمق تفضيلات الفرد أو مشاعره وهو أمر ضروري في تفسير استجاباته. فإذا كان فردان يميلان إلى لعبة رياضية معينة، فإننا نود معرفة أيهما أكثر ميلاً لهذه اللعبة، والاستناد إلى الإطار المرجعي الفردي لا يمكننا من التحقق من ذلك ؛ لأن هذا الإطار يفيد فقط في مقارنة الفرد بنفسه.

لذلك يعترض البعض على التفسيرات التى تستند إلى هذا الإطار، وبخاصة إذا حوّلنا درجات الفرد إلى معايير مئينية مثلاً، أى إذا دمجنا تفسير الدرجات استنادًا إلى كل من الإطارين الفردى والنسبى، وهذا ما يحدث بالفعل فى كثير من استبيانات الميول مثل استبيان كيودر للميول المهنية كما سنوضح بعد قليل.

فنقول مثلاً أن الفرد حصل على درجة تناظر الرتبة المئينية ٨٠ فى مقياس الميل الكتابى، أى أنه إذا واجه مواقف متعددة تماثل مواقف الحياة الواقعية، فإنه يُفضل الأنشطة الكتابية بدرجة تفوق ٨٠٪ من أقرانه، ولكن إذا حصل فردان على درجة فى الميل الكتابى تناظر الرتبة المئينية ٨٠ فإننا لا نستطيع تفسير ذلك بأنهما متساويان فى ميلهما، وذلك لأن الإطار المرجعى الذى تُنسب إليه درجة كل منهما فردى Ipsative.

ويرى بعض علماء القياس أن الدرجات التى تُستمد من الاستبيانات جبرية الاختيار Forced - Choice تكون أكثر تمثيلاً لواقع الحياة من الاستبيانات التى تكون الاختيارات فيها مفتوحة، فالفرد فى حياته يكون مضطرًا فى غالبية الأحيان إلى الاختيار بين بديلين من الأنشطة، كأن يختار أحد نشاطين ترويحيين. غير أن البعض الآخر يرى أن هذا التقييد يعمل على إحباط الفرد عند الاختيار، وبخاصة إذا كانت جميع الاختيارات إما مرغوبة أو غير مرغوبة.

أساسيات بناء استبيانات الميول المهنية:

يتطلب بناء استبيان الميول المهنية Interest Inventory اتخاذ بعض القرارات الأساسية قبل البدء فى عملية البناء نوجزها فيما يلى:

(١) ينبغى تحديد الهدف من الاستبيان وتبرير بنائه واستخدامه، حيث إننا قد ذكرنا فيما سبق أن هناك أنماطًا مختلفة للميول هى: الميول التى يعبّر عنها الفرد، والميول التى تنعكس فى سلوك الفرد، والميول التى يُستدل عليها من اختبارات التحصيل، والميول التى تُقاس بالاستبيانات. وكل من هذه الأنماط يمكن قياسها بطريقة مختلفة، مما يدعو إلى ضرورة تبرير اختيار نمط الميول التى تقاس بالاستبيانات مع ملاحظة أن جميع أدوات قياس الميول المهنية تنتمى إلى النمط الأخير.

(٢) ينبغى اتخاذ قرار يتعلق بمستويات ومجالات الميول التى يهدف الاستبيان لقياسها، حيث إن المهن والوظائف لها مستويات Levels ومجالات Areas . فالمهن Occupations مثل المحاماة والهندسة والطب والفنون والتجارة تُعد مجالات أو أنشطة متسعة يندرج تحتها العديد من الأعمال. وبعض هذه الأعمال ربما يتطلب الحصول على درجة جامعية أو تدريب متخصص، أى تكون من مستويات مهنية أو

تخصصية Professional . أما المستويات شبه المهنية Semiprofessional، فإنها تتطلب عادة دراسة جامعية أو تدريبًا متقدمًا، والمستويات المهارية Skilled تتطلب عادة تدريبًا مهنيًا، والمستويات غير المهارية لا تتطلب مهارات معينة أو ربما تتطلب قدرا محدودًا من المهارات.

لذلك ينبغي مراعاة التمييز بين مجالات ومستويات الميول المهنية عند بناء الاستبيانات من أجل اتخاذ قرارات صائبة تتعلق بميول الفرد. فالطب مثلاً يُعد مجالاً مهنيا يندرج تحته مستويات مهنية متباينة في متطلباتها التخصصية والمهارية، مثل الأعمال الفنية الطبية، والتمريض وغير ذلك. ويلاحظ أن بعض استبيانات الميول المهنية اعتمدت على تجمعات من المهن الأساسية وضُمّت مجالات ومستويات هذه المهن في معظم المقاييس المهنية التي اشتملت عليها كما في استبيان سترونج Strong، وبعضها اعتمد على قياس الميول في المجالات المهنية فقط مثل استبيان كيودر Kuder.

(٣) ينبغي اتخاذ قرار بشأن الأسلوب الذي يُتّبع في بناء الاستبيان، فقد ذكرنا فيما سبق أن هناك ثلاثة أساليب: أحدها يستند إلى أسس نظرية ومنطقية، والثاني يستند إلى بناء فقرات متجانسة، والثالث يستند إلى التحقق الإمبيريقي، وتناولنا كلًّا منها بالمناقشة. ونلاحظ أن عددًا محدودًا من استبيانات الميول المقننة استندت في بنائها على الأسلوبين الأول والثاني، بينما يبدو أن كثيرًا من الاستبيانات استندت إلى الأسلوب الثالث، حيث اختيرت فقرات الاستبيان على أساس محك إمبيريقي مناسب كما سيتضح بعد قليل.

قياس الميول باستخدام الاستبيانات المقننة:

تُعد الاستبيانات المقننة Inventories أكثر الأدوات استخدامًا في قياس الميول، فمحتوى هذه الاستبيانات يتضمن مجموعات متكاملة من الأنشطة في مجالات متعددة يعتني باختيارها. وهذا يُمكّن من التصنيف المنظم للفقرات في مجالات ميول متعددة مما يضمن عدم إغفال أي مجال ربما يختاره الفرد، وبذلك تستند درجات التفضيل إلى أسلوب كمي منظم. وتقدّر أوزان لكل فقرة بحيث تكون مميزة لمجال ميل معين عن غيره من المجالات، ومميزة لمهنة معينة عن غيرها من المهن. كما أنه يمكن تطبيق استبيانات الميول على جماعة مرجعية ممثلة تقارن في ضوئها درجة الفرد في مجالات متعددة، وكذلك مقارنة درجته بدرجات الأفراد الذين يعملون بالفعل في مهن معينة متنوعة. وبذلك يستطيع الفرد تعرّف ما إذا كان ميله مرتفعًا بدرجة كافية تشجعه على الالتحاق بمهنة معينة أو الإعداد لها.

وعلى الرغم من توافر استبيانات الميول المقننة، إلا أن عددها ليس كبيراً كما فى اختبارات الذكاء والاستعدادات والمقاييس الأخرى الشخصية. فقد أشار الكتاب السنوى السادس للقياس العقلى (Buros , 1974) إلى ٥٥ استبياناً مقنناً للميول تم نشره ومعظمها يقيس الميول المهنية، ومعظم هذه الاستبيانات من نوع الورقة والقلم، وتتطلب من الفرد تقدير أنشطة معينة أو الاختيار بين بدائل أنشطة معطاة.

وسوف نقدم فيما يلى بعض الأمثلة لهذه الاستبيانات الشائعة الاستخدام، والتى أجرى حولها كثير من الدراسات، واستند بناء كل منها إلى أحد الأساليب الثلاثة التى سبق أن أوضحناها.

(١) استبيان سترونج - كامبل للميول

The Strong - Campbell Interest Inventory (SCII)؛

يُعد هذا الاستبيان من أكثر أدوات قياس الميول شيوعاً واستخداماً، وأجرى حوله كثير من الدراسات والبحوث ما يزيد عن نصف قرن من الزمان، وقد سبق أن أوضحنا فى مستهل هذا الفصل التطور التاريخى لهذا الاستبيان . ويُعد سترونج Strong رائداً لحركة قياس الميول ، كما كان بينيه Binet رائداً لحركة القياس العقلى (Cronbach , 1965).

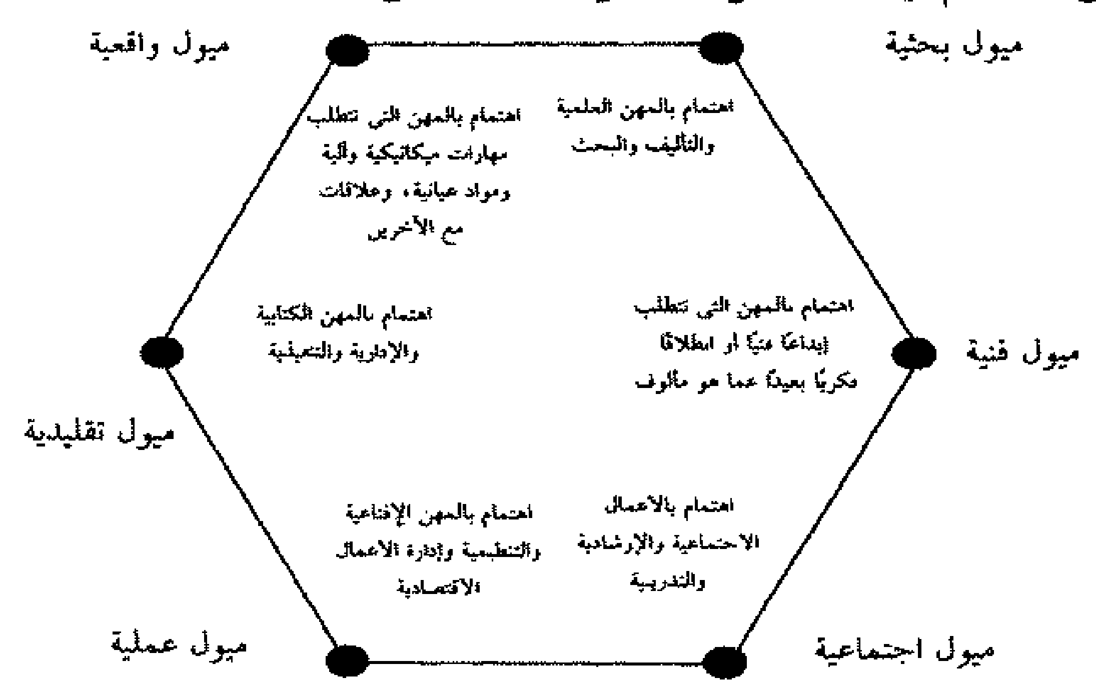
فقد قام سترونج Strong وبعض زملائه بفحص دقيق للأنشطة التى يحبها الأفراد الذين ينتمون إلى مهن مختلفة، وكذلك الأنشطة التى لا يحبونها. ولا حظوا أن أفراد المهن المختلفة يتباينون فى أنماط ميولهم، وربما يرجع ذلك إلى أن الأفراد يختارون المهن التى يميلون إليها. فالرسامون يميلون إلى اللوحات الزيتية، والذين يعملون فى حرفة النجارة يميلون إلى أعمال الأخشاب بدرجة أكبر من الذين يعملون فى حرفة الطباعة مثلاً، كما لاحظوا أن أفراد المهنة الواحدة ميولهم متماثلة ويفضلون أنشطة ترويحية واحدة، بل ويميلون إلى قراءة نوع معين من الكتب والمجلات. واستناداً إلى نتائج هذا الفحص بدأ سترونج فى إعداد استبيان لمزاوجة ميول الفرد بميول مجموعة محكة من الأفراد Criterion Group الذين ينتمون إلى مهن مختلفة يحبونها وناجحون فيها.

لذلك استند سترونج فى بنائه للاستبيان إلى أسلوب التحقق الإمبيريقى Empirical Keying ، أى الأسلوب الثالث الذى سبق أن أوضحناه. وقد ظهرت الطبعة الأولى من هذا الاستبيان الذى أطلق عليه استبيان سترونج للميول المهنية (SVIB) عام ١٩٢٧ ، وهذه الطبعة كانت خاصة بالذكور، وصدرت الطبعة الأولى من الاستبيان الخاصة بالإناث عام ١٩٣٣. وتمت مراجعة الطبعة الخاصة بالذكور عامى

١٩٣٨ ، ١٩٦٦، والطبعة الخاصة بالإناث عامى ١٩٤٦ ، ١٩٦٩، ولم تنل الطبعة الخاصة بالإناث اهتمامًا مماثلاً لما نالته الطبعة الخاصة بالذكور، وذلك لأن الإناث فى ذلك الوقت لم يكونوا جزءًا فاعلاً من القوة البشرية العاملة.

وقد قام كامبل Campbell وهو أحد علماء النفس بمراجعة الاستبيان ومحاولة التغلب على أوجه القصور فيه، ودمج فقرات الطبعتين معًا فى طبعة واحدة محايدة يمكن تطبيقها على كل من الذكور والإناث، وأطلق عليه استبيان سترونج - كامبل للميول (SCII) عام ١٩٧٤،وصدرت طبعة حديثة من الاستبيان عام ١٩٨١،وبذلك تغلب على مشكلة التحيز للذكور. وقد بينت البحوث أن الميول المهنية لا تستقر قبل عمر ١٧ أو ١٨ عامًا أو ربما بعد ذلك.

لذلك فإن هذا الاستبيان يُطبق عادة على الراشدين، ولكن يمكن تطبيقه على الشباب صغار السن حيث إن صياغة فقراته بسيطة، وهو غير موقوت، ولكن تطبيقه يستغرق ساعة واحدة تقريبًا. ونظرًا لأنه كان يُعاب على الاستبيان عدم استناده إلى إطار نظرى يفسر تفضيلات الأفراد الذين ينتمون إلى مهن معينة، فقد حاول كامبل الاسترشاد بالنظام التصنيفى للميول المهنية للأفراد الذى اقترحه هولاند Holland عام ١٩٧٥ . ويرى هولاند أن الميول تُعبر عن شخصية الفرد المميزة، واقترح تصنيف الميول المهنية فى ستة أقسام يوضحها الشكل التخطيطى (١١ - ٤) التالى (Holland , 1973):



شكل (١١ - ٤) يوضح تصنيف هولاند للميول المهنية

وقد أوضحت الدراسات أن هذا النظام التصنيفى أفضل من غيره من الأنظمة التصنيفية للمهن فى وصف الأنشطة المتعلقة بالأعمال المختلفة، والمتطلبات العامة للتدريب، والحوافز المهنية. ولعل اتفاق هذا التصنيف مع أنماط الميول المهنية التى اهتم بها سترونج فى استبيانه الأصلى هو الذى شجع كامبل على الاسترشاد به فى الاستبيان.

ويشتمل الاستبيان فى طبعته الحديثة على ٣٢٥ فقرة تم تجميعها فى سبعة أقسام من حيث المحتوى كالتالى (Campbell & Holland , 1981) :

(١) المهن (١٣١ فقرة) ، مثل : المحامى الجنائى، الكاتب الحر، الممثل، فنى أشعة.

(٢) المجالات الدراسية (٣٦ فقرة) ، مثل : الأدب، الاقتصاد، الفنون.

(٣) الأنشطة (٥١ فقرة) ، مثل : تجبير العظام، الطهى، مشاهدة فيلم طبى.

(٤) الأنشطة الترويحية (٣٩ فقرة) ، مثل : الألعاب الرياضية، الاستماع إلى الموسيقى الكلاسيكية، صيد السمك.

(٥) أنماط من الناس (٣٤ فقرة) ، مثل : كبار السن، أفراد انفعاليين، مغايرين.

(٦) التفضيل بين نشاطين (٣٠ فقرة) ، مثل : التفضيل بين العمل الحكومى والعمل الحر، أو بين قراءة كتاب ومشاهدة فيلم.

(٧) الخصائص المميزة للفرد (١٤ فقرة) ، مثل : استطاعة الفرد تصميم إعلان متميز، تكوين صداقات بسهولة.

ويجيب الفرد بوضع علامة مقابلة لكل فقرة توضح ما إذا كان يحب النشاط المحدد، أو لا يحبه ، أو لا يهتم به.

وفيما يلى مثال من كل قسم من الأقسام السبعة السابقة :

	أحب	لا أهتم	لا أحب
مهندس معمارى	...	...	...
الاقتصاد	...	...	...
مشاهدة فيلم طبى	...	...	...
صيد السمك	...	...	...
الرحالة	...	...	...

	أفضل	محايد	لا أفضل
أشاهد مباراة كرة قدم مقابل الذهاب إلى متحف	...	...	...

	نعم	لا
يمكنني الترويح عن الآخرين	...	...

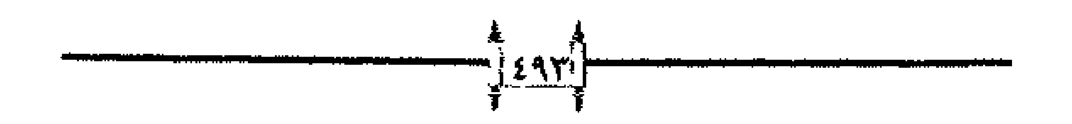
ومن الجدير بالذكر أن طبعة عام ١٩٨١ أضافت عددًا من المهن حيث بلغ عددها ١٦٢ مهنة، كما أن الجماعة المرجعية المحكة لكل مهنة أصبحت ممثلة بعينة قومية، وروعيت بعض الاحتياطات التي تُيسر تفسير الدرجات. فقد كان يعاب على الاستبيان أن أفراد المجموعات المحكة أكبر سنًا من الأفراد المبتدئين في المهنة. لذلك روعي في هذه الطبعة المقارنة بين صغار السن والكبار في كل مجموعة محكة لتحديد ما إذا كانت الميول والقيم للمبتدئين تختلف عن ميول وقيم الذين قضوا في العمل عدداً من السنوات. كذلك اشتملت هذه الطبعة على معايير خاصة بالذكور والإناث لكل من هذه المهن. ويتم تصحيح الاستبيان وتقدير درجاته وإعداد البروفيل باستخدام حاسوب مركزي، وتحوِّل درجات المقاييس الفرعية التي يشتمل عليها الاستبيان إلى درجا ـ معيارية.

ويتضمن تقرير النتائج ما يلي:

(١) **ستة توجُّهات مهنية عامة** : تصف النمط أو التوجُّه العام للميول المهنية للفرد كما هو موضح بالشكل التخطيطي (١١ ـ ٤)،فالتوجُّهات المتقاربة معًا في الشكل التخطيطي المذكور تكون أكثر تماثلاً من التوجُّهات المتباعدة، وتقدر درجة معيارية لكل نمط أو توجه.

(٢) **الميول الأساسية** : تصف مجموعات من الأنشطة الأكثر تحديدًا والمرتبطة بالتوجهات المهنية العامة الستة، وعدد مقاييسها الفرعية ٢٣. فمثلاً ترتبط أنشطة التدريس، والخدمة الاجتماعية، والرياضيات، والفنون الشعبية بالتوجُّه أو الميول الاجتماعية. وتتوسط الميول الأساسية بين التوجُّهات المهنية والميول المهنية الأكثر نوعية.

فهي تصف تجمُّعات أو مجالات من الأنشطة التي يحبها الفرد ، وهذه المجالات ربما تشترك فيها كثير من المهن النوعية.



(٣) **الميول المهنية** : تصف مدى تماثل نمط ميول الفرد مع ميول الأفراد الناجحين في مهن معينة عددها ١٦٢ مهنة، مثل : مزارع ، أو ضابط طيار، أو مسئول إعلانات، أو فنان، أو معلم رياضيات وعلوم. ويحصل الفرد على درجة معيارية واحدة لكل مهنة.

(٤) **ميول غير مهنية** : يوجد مقياسان فرعيان للميول غير المهنية Nonoccupational هما:مقياس التوجه الأكاديمي، ومقياس الانطواء ـ الانبساط. ويقيس الأول مدى حب الفرد للأنشطة العقلية، ويقيس الثاني مدى حب الفرد للتفاعل مع الآخرين والعمل معهم في مقابل حبه للعمل منفرداً. وقد استند في بناء هذين المقياسين على محكات غير مهنية. فقد قورن في المقياس الأول بين النسبة المئوية المناظرة لاستجابات طلاب الكليات المتفوقين في دراستهم والنسبة المئوية المناظرة للطلاب الضعفاء. وقورن في المقياس الثاني بين الطلاب الذين أمكن تحديدهم بأنهم منطوون أو منبسطون باستخدام استبيان الشخصية متعدد الأوجه (MMPI). وقد بينت الدراسات أن هناك فروقًا كبيرة بين متوسطات الدرجات في كل من المقياسين في الاتجاه المطلوب بين المجموعات التي تنتمي إلى مهن مختلفة.

ويُستخدم هذا الاستبيان في معظم الأحيان في أغراض الإرشاد حيث يتطلب متخصصًا لديه خبرة في تفسير نتائجه تفسيراً مستنيراً، وبخاصة أن نتائجه تكون في شكل بروفيل أو صفحة نفسية متعددة الجوانب يُعدّها الحاسوب. ولتيسير تفسير النتائج أوكل إلى عدد من المؤسسات المتخصصة بالولايات المتحدة الأمريكية، مثل نظم الحاسوب القومية National Computer Systems، وشركة Mincomp إعداد تقارير متعددة الصفحات تتعلق بتفسير نتائج الاستبيان المتمثلة في البروفيل وتقديم المقترحات للعميل والمرشد.

وتكون مهمة المرشد عندئذ هي معاونة العميل في تحديد الاتساق بين الدرجات في مختلف أقسام المقاييس، أو عدم الاتساق بين درجات كل من مقياس الميول المهنية، ومقياس التوجهات المهنية العامة. ولعل من الأفضل استخدام هذا الاستبيان إلى جانب المعلومات التي يكشف فيها الفرد عن ميوله لفظيًا (الميول التي يعبّر عنها الفرد) ، وكذلك مؤشرات عن قدراته باستخدام اختبارات القدرات العقلية التي اعتنى بتقنينها.

(٢) استبيانات كيودر للميول The Kuder Interest Inventories:

أعدّ كيودر Kuder مجموعة من الاستبيانات التي تقيس الميول بتوجّهات تختلف عن التوجهات التي استند إليها استبيان سترونج ـ كامبل ، فقد أعد كيودر استبيان

التفضيلات المهنية Kuder Vocational Preference Record عام ١٩٣٩، وظهرت طبعته الثانية عام ١٩٦٤، وأطلق عليه استبيان مسح الميول العامة

. Kuder Form E General Interest Survey (GIS)

ويُطبق على طلاب المدارس الثانوية والكليات والراشدين من الجنسين.

واستند في بنائه إلى انتقاء فقرات متجانسة Homogenous Keying (الأسلوب الثاني) حيث تم تجميع فقرات كل مجال عريض من مجالات الميول، مثل الميول الأدبية، أو الميول العلمية، وغيرهما في مقياس خاص به، وتطبيق الفقرات على عينة تجريبية من الأفراد، وإيجاد ارتباط درجات كل فقرة بالدرجات الكلية في المقياس الذى تنتمي إليه، واستبعاد الفقرات التي تكون ارتباطاتها منخفضة.

وبذلك أمكن الحصول على عشرة مقاييس متسقة اتساقًا داخليًا تتعلق بالميل الخلوى، والميكانيكي، والحسابي، والعلمي، والإقناعي، والفني، والأدبي، والموسيقي، والخدمة الاجتماعية، والكتابي . وتحوّل درجات الفرد في كل من هذه المقاييس إلى رتب مئينية بحيث يمكن مقارنة نمط ميوله بنمط ميول الأفراد الذين يحبون المجال الذى يعملون فيه. واستخدم كيودر فقرات جبرية الاختيار يشتمل كل منها على ثلاثة بدائل أنشطة Forced - Choice Triad Items، حيث يقدّم ثلاثة أنشطة يختار الفرد من بينها النشاط الذى يحبه بدرجة أكبر، والنشاط الذى يحبه بدرجة أقل، مثل :

(أ) أكتب خطابًا لأحد الأصدقاء.

(ب) أقوم بإعداد ترتيبات حفل.

(جـ) أقوم بتحصيل الفواتير المتأخرة السداد من العملاء.

وهذا يختلف عن استبيان سترونج - كامبل الذى استخدم فقرات تتطلب تحديد ما إذا كان الفرد يحب نشاطًا معينًا أم لا يحبه، أى أن الاختيار هنا مطلق، بينما في استبيان كيودر يكون التفضيل نسبيًا، وبذلك نحصل في النهاية على تراتيب للتفضيلات، وهو ما أطلقنا عليه فيما سبق القياس الذى يُنسب إلى إطار مرجعي فردى Ipsative Measurement، أى أن درجة الفرد في أحد المقاييس التي يشتمل عليها الاستبيان تتأثر بدرجاته في المقاييس الأخرى. فإجراءات التصحيح وتقدير الدرجات في هذه الحالة ترتّب الميول من حيث شدتها النسبية داخل الفرد نفسه، ولا تُعطى أى مؤشرات عن الشدة المطلقة لميوله، فالفرد الذى لا يهتم بعامة بمجالات معينة ولكن لديه ميل قوى نحو مجال واحد سوف يحصل على درجة مرتفعة في هذا المجال.

لذلك فإنه على الرغم من ميزات انتقاء فقرات الاستبيان على أساس منطقى واتساق داخلى، إلا أنه لا ينبغى إغفال المشكلات المتعلقة بتفسير الدرجات فى هذه الحالة. كما أن استخدام البدائل الثلاثة فى الاختيار تؤدى أيضًا إلى مشكلات أخرى، فأحيانًا ربما لا يُفضل الفرد الأنشطة الثلاثة التى تُمثلها كل من هذه البدائل، وأحيانًا أخرى ربما يُفضل هذه الأنشطة جميعها، مما يؤدى إلى إرباك الفرد واستغراق وقت كبير فى الاستجابة.

وقد أعد كيودر بعد ذلك استبيانًا آخر أطلق عليه «استبيان كيودر لمسح الميول المهنية The Kuder Occupational Interest Survey (KOIS) فى أواخر الستينيات وتمت مراجعته وأعيدت طباعته عام ١٩٨٥ (Kuder & Diamond , 1985)، ويُستخدم مع الأفراد من سن أصغر حيث توجد معايير للسن من ٦ - ١٢. وقد استند فى بنائه إلى أسلوب التحقق الإمبيريقى مثل استبيان سترونج - كامبل بدلاً من أسلوب انتقاء الفقرات المتجانسة الذى استند إليه فى استبيانه السابق. ولكنه استبقى على نظام الفقرات جبرية الاختيار ذات البدائل الثلاثة كما سبق.

ويختلف هذا الاستبيان عن استبيان سترونج - كامبل فى أن الدرجة فى كل مقياس يقيس ميلاً مهنيًا معينًا تُمثل بالارتباط بين نمط ميل الفرد ونمط أفراد جماعة مرجعية فى مهنة معينة، فقد أوجد كيودر معامل الاقتران بين استجابات الفرد لفقرات الاستبيان ونسبة عدد الأفراد فى مجموعة مهنية معينة الذين استجابوا للفقرات. فكلما زادت قيمة معامل الاقتران دل ذلك على اقتران أو تماثل درجة الفرد بنمط ميول هذه المجموعة.

وقد تم تقنين الاستبيان على عينة قوامها ١٢٠٠٠ طالب بالسنة الأولى بالكليات المختلفة فى الولايات المتحدة الأمريكية، واشتمل الاستبيان فى طبعته الأخيرة على ٧٧ مقياسًا للميول المهنية، ٤٩ مقياسًا للميول نحو تخصصات أكاديمية للذكور، ٥٧ مقياسًا للميول المهنية، ٢٧ مقياسًا للميول نحو تخصصات أكاديمية للإناث، والاستبيان غير موقوت، ويستغرق تطبيقه ساعة واحدة تقريبًا. ولتصحيح الاستبيان وتقدير درجاته ينبغى إعادة الاستبيان بعد استيفائه إلى الناشر لإجراء ذلك باستخدام الحاسوب.

وقد بيَّنت الدراسات أن درجات هذا الاستبيان أكثر استقرارًا من درجات استبيان سترونج - كامبل فى أغراض إرشاد الطلاب للمهن غير المتخصصة، وتراوحت قيم ثبات درجات استبيانات كيودر بين ٠,٧٠ ، ٠,٩٠ ، وعلى الرغم من أن الثبات المتعلق بتقديرات الاتساق الداخلى يشير إلى اتساق بروفيل الميول فى مدى زمنى قصير، إلا أنه ربما لا نستطيع توقع اتساقها على مدى زمنى طويل ما لم تتبلور ميول الفرد.

لذلك يلاحظ تذبذب بروفيل ميول الطلاب الذين تقل أعمارهم عن ١٧ أو ١٨ عام في كل من استبيان كيودر واستبيان سترونج (Crites , 1969) . وتستقر الميول نسبيًا لدى طلاب الكليات ، ويزداد استقرارها في نهاية المرحلة الجامعية (Campbell , 1966) .

(٣) استبيان لى ـ ثروب للميول المهنية

Lee - Thrope Interest Inventory :

على الرغم من أن هذا الاستبيان يقتصر استخدامه في البحوث نظرًا لعدم وجود أدلة كافية تتعلق بثبات درجاته وصدقه، إلا أننا آثرنا أن نشير إليه بإيجاز لأنه استند في بنائه إلى الأسلوب المنطقي (الأسلوب الأول) وبذلك نكون قد قدّمنا أمثلة لاستبيانات استندت في بنائها إلى كل من الأساليب الثلاثة التي سبق أن أوضحناها. ومن الجدير بالذكر أن هذا الأسلوب المنطقي محدود الاستخدام في مجال بناء المقاييس بعامة.

وقد أعد هذا الاستبيان لى Lee، وثروب Thrope وأعيد طباعته بعد مراجعته عام ١٩٥٦ ، ويتضمن مستويين: أحدهما يناسب طلاب المرحلة الثانوية، والآخر مستوى متقدم يناسب الراشدين، وهو استبيان غير موقوت ويستغرق تطبيقه ٤٥ دقيقة تقريبًا.

وقد قام المؤلفان بفحص دليل عناوين المهن الذى أعدّه مكتب خدمات التوظيف بالولايات المتحدة الأمريكية، وقاموا بتصنيف المهن في ستة أقسام عريضة أو مجالات ميول وهي : شخصي ـ اجتماعي ، طبيعي ، ميكانيكي، تجاري، فنون، عا م. وأعدوا استبيانًا يمكن باستخدامه الحصول على درجات يمكن تصنيفها في مجموعات ثلاث هي: مجالات الميول، وأنواع الميول، مثل الميل اللفظي والميل اليدوى والميل الحسابي، ومستوى الميول. وقاموا ببناء فقرات لكل من هذه الأقسام الستة تمثل المستويات الثلاثة للميول، أي ميل مرتفع ومنخفض ومتوسط. واشتمل الاستبيان على ١٢٠ زوجًا من الفقرات من نوع الاختيار الجبرى، وكما ذكرنا فإن هذا الاستبيان يحتاج إلى دراسات للتحقق من خصائصه السيكومترية.

وبالطبع توجد استبيانات أخرى كثيرة لقياس الميول، وسوف نشير إلى بعض منها فقط فيما يلي:

(١) استبيان جاكسون لمسح الميول المهنية

The Jackson Vocational Interest Survey (JVIS) :

ويُعد استبيانًا حديثًا نسبيًّا حيث صدرت طبعته الثانية عام ١٩٨٢ ، وأشار مؤلفه (Jackson , 1982) إلى أنه أفاد من التطورات التقنية لإحداث تكامل بين الاختبارات والأساليب المتقدمة في بنائها.

واستند في بنائه إلى تصنيف هولاند الموضح بشكل (١١ ـ ٤) . ويشتمل الاستبيان على فقرات من نوع الاختيار الجبري، حيث يختار الفرد بين بديلين يتعلقان بأدوار عمل Work roles ، أو أساليب عمل Work Styles ، ونتيجة ضم البيانات يمكن الحصول على درجات معيارية لعدد يبلغ ٣٤ مقياسًا يتعلق ٢٦ منها بأدوار العمل، و٨ بأساليب العمل، وقد صممت هذه المقاييس لمقارنة الميول المهنية للفرد بميول جماعات مرجعية من الذكور والإناث، وإعطاء صورة ممثلة للتوجهات المهنية العامة للفرد في ضوء تصنيف هولاند، ويعد تقرير بروفيل النتائج لكل فرد باستخدام الحاسوب. ومما يميز هذا الاستبيان أن مؤلفه قدم العديد من الأدلة عن صدق التكوين الفرضي، وبذلك ربما يصبح هذا الاستبيان من الأدوات المهمة في قياس الميول المهنية.

(٢) استبيان مينيسوتا للميول المهنية

The Minnesota Vocational Interest Inventory (MVII) :

أعدَّ هذا الاستبيان Clark ونُشر عام ١٩٦٥ويمكن باستخدامه الحصول على معلومات منظمة تتعلق بأنماط الميول المتعلقة بالذكور في المهن غير التخصصية أي الحرف التي تتطلب مهارة. وقد أُعد هذا الاستبيان في ضوء نتائج البحوث التي أُجريت لتوسيع نطاق استبيان سترونج واستبيان كيودر، حيث إنهما لا يناسبان المهن ذات المستويات الأدنى. وهذا الاستبيان يناسب الذكور الذين تبلغ أعمارهم ١٥ عامًا أو أكثر، ويستغرق تطبيقه ٤٥ دقيقة. ويمكن الحصول على درجات تتعلق بحرف مختلفة يبلغ عددها ٢١ حرفة أو صنعة مثل: فني طباعة، خبّاز، نقّاش، كهربائي وغير ذلك. واستند في بنائه إلى الأسلوب الإمبيريقي ، وذلك بمقارنة استجابات الأفراد الذكور الذين ينتمون إلى كل من هذه الحرف بالعاملين الحرفيين بعامة.

بالإضافة إلى أن الاستبيان يتضمن تسعة مقاييس للميول العامة الميكانيكية، والخدمات الصحية، والأعمال المكتبية، والإلكترونيات، والخدمات الغذائية، وأعمال النجارة، والمبيعات، والأعمال الخلوية وغيرها.

وهذه المقاييس تشتمل على مجموعات من الفقرات المترابطة ، وتعتمد معايير الاختبار على الدرجات التائية حيث تُعبِّر الدرجة ٥٠ عن متوسط درجات المجموعة المحكة في كل من المقاييس المتعلقة بالحرف المختلفة. وتتراوح قيم ثبات درجات هذا الاستبيان بطريقة الإعادة بمدة زمنية فاصلة تبلغ شهرًا بين ٦٢, ٠ ، ٨٨, ٠ .

(٣) استبيان أوهايو لمسح الميول المهنية

: The Ohio Vocational Interest Survey (OVIS)

أعدَّ هذا الاستبيان دكوستا D'Costa ووانيفوردنر Winefordnerبجامعة أوهايو عام ١٩٦٦، وتم تقنينه عام ١٩٦٩، وقد صُمِّم لكي يُستخدم في إرشاد طلاب المرحلة الثانوية في اتخاذ قراراتهم المهنية، ويستغرق تطبيقه ما بين ساعة وساعة ونصف. وقد استند الاستبيان إلى دليل عناوين المهن الذي أشرنا إليه سابقًا لتصنيف المهن. وأعد المؤلفان في ضوء ذلك ٢٤ مقياسًا فرعيًا للميول عيّن لكل منها رمز من ثلاثة أرقام يشير الرقم الأول إلى «بيانات Data» ، والرقم الثاني إلى أفراد People، والرقم الثالث إلى «أشياء Things» كما وضحه الدليل المذكور في شكل مكعب . فمثلاً العمل المتعلق بالأعمال الفنية الطبية يرمز له بالرمز(٢٢٢) حيث تشير هذه الأرقام إلى أن العمل يتطلب التعامل مع البيانات والأفراد والأشياء بدرجة كبيرة، نظرًا لأن الأرقام المناظرة لكل بُعد من الأبعاد الثلاثة للمكعب (بيانات، وأفراد،وأشياء) تتراوح بين صفر (لا يوجد تعامل)، ٢ (تعامل بدرجة كبيرة).

ويشتمل كل من هذه المقاييس على ١١ فقرة تشير إلى أنشطة يستجيب لها الفرد. ولكن تظهر الفقرات منفردة، وليست ثلاثيات كما في استبيان كيودر، ولكل فقرة ميزان متدرج مكون من خمس نقاط (أحب جدًا، أحب، محايد، لا أحب، لا أحب على الإطلاق). ومن أمثلة هذه الفقرات: (أزرع الخضروات، أصمم رسومًا للأطفال، أكتب برامج للحاسوب، أصلح المناضد والمقاعد)، وتصحيح الاستبيان مماثل لتصحيح استبيان كيودر.

وتعكس نتائج الاستبيان الميول النسبية للفرد في مختلف الحرف، كما تعكس ميله لكل حرفة مقارنة بميل العاملين فيها من الجماعة المرجعية مما يساعد في إلحاق الأفراد ببرامج التوعية المهنية. وقد استُمدت معايير الاختبار من عينة قومية تشتمل على ٤٥٠٠٠٠ طالب في المرحلة الثانوية بالولايات المتحدة الأمريكية.

(٤) استبيان جايست المصور للميول

: The Geist Picture Interest Inventory (GPII)

يختلف هذا الاستبيان عن الاستبيانات السابقة في أنه لا يعتمد على الجوانب اللفظية، وإنما يعتمد على الصور، وقد نُشر أول مرة عام ١٩٥٨. والاستبيان يناسب الأفراد الذين لديهم صعوبات في القراءة، ويقتصر استخدامه على الذكور، ويمكن تطبيقه فرديًا أو جماعيًا. ويشتمل على ٤٤ فقرة مصورة يتضمن كل منها بدائل ثلاثية من الصور المتعلقة بمختلف المهن والهوايات. ويلي كل فقرة ثلاثية سؤال يجيب عنه الفرد باختيار البديل الذي يفضله.

ويذكر جايست (Geist , 1959) مؤلف الاستبيان، أنه يمكن استخدامه في المواقف التي تتطلب تخطيطًا واختيارًا مهنيًّا، ودراسة أنماط القوى الدافعية المرتبطة بالميول، وكذلك دراسة الميول ذاتها. ويقترح أنه يمكن الإفادة من الاستبيان في معاونة المعلمين اختيار أنشطة تعليمية مفيدة، وربما أيضًا في دراسة أسباب اختيار الأطفال هوايات محددة، واختيار الكبار مهنًا معينة. ويمكن باستخدام الاستبيان الحصول على ١١ درجة للميول، ٨ درجات للدافعية. وتشير درجات الميول إلى: الميل الإقناعي، والكتابي، والميكانيكي، والدرامي، والموسيقي، والعلمي، والخلوي، والأدبي، والحسابي، والفني، والخدمة الاجتماعية. أما درجات الدافعية فتشير إلى : الأسرة، المركز الاجتماعي، العائد المادي، الدافع الذاتي، الشخصية، الوسط البيئي، الخبرة السابقة، لا أعرف.

وقد أُجري تقنين الاستبيان على ١٢٠٠٠ فرد من جميع المستويات العمرية في الولايات المتحدة الأمريكية. وهاواي، وبورتوريكو. وتعتمد المعايير على الدرجات التائية لكل مجال من مجالات الميول المذكورة، وتُمثَّل هذه الدرجات في بروفيل لكل فرد يوضح مستويات الميول وقوة الدافعية. وأوضحت الدراسات أن قيمة معامل ثبات درجات الاستبيان بطريقة الإعادة بمدة زمنية فاصلة بلغت ستة أشهر لعينات من طلاب المدارس الثانوية والكليات قد تراوحت بين ٠,٣٧ ، ٠,٩٤ لجميع المقاييس الفرعية، أما صدق الاستبيان فيتطلب مزيدًا من الدراسات.

تعقيب عام على استبيانات الميول :

اتضح لنا من الأمثلة السابقة لاستبيانات الميول أنها استندت في بنائها إلى أساليب متباينة واشتملت على فقرات مختلفة، واعتمد بعضها مثل استبيان كيودر على أنواع الفقرات التي تتطلب من الفرد الاختيار الجبري من بين بديلين أو ثلاثة، أي أن الفرد يقوم بترتيب البدائل بحسب درجة تفضيله لها، وهذا يعني أنه كلما قَبِل الفرد مجال ميل معين ينبغي عليه رفض مجال ميل آخر، إذ لا يمكن أن يقبل جميع المجالات.

واعتمد البعض الآخر مثل اختبار سترونج على فقرات تتطلب استجابة غير مُقيَّدة، وذلك بأن يوضح الفرد حبه لنشاط معين أو عدم حبه له. غير أن هذا النوع من الفقرات أكثر تأثرًا بالاتجاهات العقلية في الاستجابات من النوع السابق على الرغم من أن تفسير درجات الاستبيان الذي يشتمل على فقرات جبرية الاختيار يستند إلى الإطار المرجعي الفردي Ipsative Scores، مما يؤدي إلى صعوبات في تفسير هذه الدرجات كما سبق أن أوضحنا.

وبعض الاستبيانات مثل استبيان سترونج، واستبيان مينيسوتا، واستبيان كيودر في طبعتها الحديثة استندت في بنائها إلى أسلوب التحقق الإمبيريقي، حيث استخدم كل من الاستبيان الأول والثاني الناس بعامة كمجموعة محكة.

بينما قارن الاستبيان الثالث الأفراد الذين ينتمون إلى مهنة معينة بغيرهم في المهن الأخرى. أما استبيان سترونج لمسح الميول العامة فقد استند إلى أسلوب انتقاء فقرات متجانسة، واستند استبيان لي - ثروب للميول المهنية في بنائه إلى الأسلوب المنطقي. وبعض الاستبيانات مثل استبيان سترونج واستبيان كيودر اشتملا على أنشطة وميول تتعلق بالمهن والحرف التخصصية.

بينما تركز استبيان مينيسوتا على الأنشطة والميول المتعلقة بالمهن غير التخصصية أو الحرف التي تتطلب مهارة، أما استبيان أوهايو لمسح الميول المهنية فقد اعتمد على دليل عناوين المهن، واشتمل على العديد من الأنشطة التي ربما يكون لدى طلاب المرحلة الثانوية معلومات أو خبرة محدودة حولها. وجميع استبيانات الميول غير موقوته، ويستغرق تطبيقها ساعة واحدة تقريبًا بطريقة جماعية، ويمكن تطبيقها فرديًا. وكثير من هذه الاستبيانات وبخاصة الاستبيانين الرئيسين لسترونج وكيودر يتم تصحيحهما وتقدير درجاتهما باستخدام حاسوب مركزي، وتُعاد النتائج للجهة المُرسلة.

وتتباين استبيانات الميول في مدى سهولة تصحيحها وتقدير درجاتها، حيث يصعب إجراء ذلك يدويًا في بعض هذه الاستبيانات وبخاصة استبياني سترونج وكيودر.

وفيما يتعلق بصدق استبيانات الميول وثبات درجاتها فإن استبيان سترونج يُعد أكثر الاستبيانات التي نالت دراسات صدق وثبات متعددة، يليه استبيان كيودر، وأقلها استبيان جايست المصور للميول. أما فيما يتعلق بتفسير درجات هذه الاستبيانات فقد اتضح لنا أن بعضًا منها يستند إلى الإطار المرجعي الفردي Ipsative نظرًا لأنها تشتمل على فقرات جبرية الاختيار مثل استبياني كيودر ومينيسوتا، والبعض الآخر يستند إلى إطار مرجعي نسبي أي إلى معايير الجماعة؛ وعددًا محدودًا استند إلى كل من الإطارين.

ولعل استبيان كيودر هو الاستبيان الوحيد الذي اشتمل على مفاتيح لكشف عدم فهم التعليمات أو الفقرات، وعدم تحرّي الدقة والأمانة في الاستجابات.

وعلى الرغم من هذا الاختلاف في أساليب البناء، وتقدير الدرجات، وتقدير نتائج مقاييس فرعية متعددة بتسميات مختلفة، إلا أن كول وهانسون (Cole & Hanson , 1961) قد أوضحا أن معظم هذه الاستبيانات وما تشتمل عليها

من مقاييس فرعية يمكن تصنيفها بسهولة فى إطار نموذج هولاند المبين بشكل (١١ـ٤) الذى يشتمل على ستة عوامل تتعلق بالشخصية الإنسانية والمهن المختلفة . ويرى جيلفورد Guilford ومعاونوه أن استبيانات الميول ربما يمكن استخدامها كمقاييس للشخصية .

واقترحوا أنه لا مجال للشك فى أن الميول ينبغى أن ترتبط بالدوافع والطموحات والحاجة نظراً لأنها لا تتكون أو تنمو فى فراغ، ولعل الدراسات الحديثة المتعلقة بمجال النمو المهنى أكدت صدق هذا الاقتراح (Super : 1972 ; Holland 1973 ; Tayler , 1983).

استخدامات استبيانات الميول :

يمكن استخدام استبيانات الميول استخدامات متعددة فى المجالات التربوية والمهنية، غير أن الإفادة منها فى هذه المجالات يعتمد على الغرض من استخدامها والقرارات التى تتخذ فى ضوئها. فاستبيانات الميول تُعد أحد أدوات قياس خصائص الأفراد وسلوكهم المميز، لذلك ينبغى مراعاة حدود استخدام هذه الاستبيانات وعلاقتها بأدوات قياس الجوانب المعرفية المتعلقة بالسلوك الأقصى للأفراد. وسوف توضح فيما يلى أهم أغراض استخدام هذه الاستبيانات ومدى الإفادة منها فى كل من هذه الأغراض:

أغراض التوجيه التربوى والمهنى : تُستخدم استبيانات الميول بكثرة في أغراض التوجيه والإرشاد التربوى والمهنى إلى جانب اختبارات التحصيل والاستعدادات العقلية. فإرشاد أو توجيه الفرد إلى الدراسة أو المهنة المناسبة له يتطلب بيانات متعددة لا نحصل عليها من استبيانات الميول بمفردها. غير أن هذه الاستبيانات إذا استُخدمت استخدامًا واعيًا يمكن أن تزود المرشد التربوى أو المهنى بمعلومات أساسية تتعلق بميول الفرد أو تفضيلاته الأكاديمية والمهنية.

ومع هذا فإن هناك أدلة متعددة تشير إلى أن ما يقرره الأفراد من تفضيلات مهنية ربما يكون غير واقعى، وبخاصة فى بدء المرحلة الثانوية وربما بعد ذلك، نظراً لأن هؤلاء لا يكونون على دراية تامة بالأنشطة المتعلقة بالمهن المختلفة ومتطلباتها، بل إن كثيراً من الشباب ربما لا يكون لديهم خبرة كافية للاختيار من بين مدى واسع من الأنشطة التى تتضمنها الاستبيانات . وتزداد حدة المشكلة إذا استُخدمت هذه الاستبيانات فى المرحلة المتوسطة حيث يجد بعض الطلاب صعوبات فى فهم المصطلحات المهنية التى يقررون فى ضوئها اختيارهم المناسب.

لذلك يمكن أن يستخدم المرشد نتائج هذه الاستبيانات فى إلقاء الضوء على بعض المشكلات التى يمكن أن يُقدّم فيها العون للفرد، مثل عدم اهتمامه بالعمل

الجماعى، أو العلاقات الشخصية، أو الصعوبات الدراسية وغير ذلك. كما يمكنه استخدامها فى تعرّف مدى اتساق نمط الميول التى عبّر عنها الفرد فى استجاباته، وما إذا كان البرنامج الإرشادى يناسب بروفيل ميوله واستعداداته.

وينبغى عدم اتخاذ قرارات فاصلة استناداً إلى نتائج استبيانات الميول. إذ لا ينبغى أن تكون هذه النتائج هى المحك الأساسى الذى يُستند إليه إرشاد الفرد إلى تخصص أكاديمى معين أو مهنة محددة. لذلك يجب على المرشد عند تفسيره لنتائج استبيانات الميول أن يُوضح للفرد الفرق بين الميول والاستعدادات.

فكثير من الأفراد يستنتجون أن لديهم القدرات أو الاستعدادات التى تؤهلهم للنجاح فى مهنة معينة وذلك لأنهم يميلون إليها، فى حين أن الاستبيان لا يوضح سوى أن ميول الفرد تماثل ميول جماعة مرجعية من العاملين فى هذه المهنة، وقد سبق ذكرنا أن الارتباط بين الميول والاستعدادات يميل إلى الانخفاض، مما يدل على أن الأفراد فى غالبية الأحيان لا يمتلكون القدرات أو الاستعدادات التى تتطلبها الأنشطة المعينة التى يميلون إليها ، وربما يكون العكس صحيحًا. لذلك ينبغى تفسير نتائج استبيانات الميول فى إطار معلومات متكاملة عن قدرات الفرد واستعداداته الخاصة ودافعيته وتحصيله وسمات شخصيته.

أغراض الانتقاء والتسكين: يختلف استخدام استبيانات الميول فى أغراض الانتقاء والتسكين عنه فى أغراض الإرشاد والتوجيه، ففى أغراض الانتقاء يترتب على نتائج الاستبيانات قرارات تمس مستقبل الفرد مباشرة.

فالفرد فى هذه الحالة يبغى الالتحاق بمهنة أو وظيفة شاغرة، لذلك فإنه ربما يلجأ إلى تزييف استجاباته أو جعلها متحيزة للمهنة أو الوظيفة التى تقدّم إليها، فقد بينت معظم الدراسات أن الاتجاهات العقلية فى الاستجابات Response Sets التى سبق أن أوضحناها ربما تلعب دوراً كبيراً فى هذه المواقف الانتقائية ويصعب التغلب عليها (Kuder , 1966). فهناك أفراد يحاولون إعطاء انطباع جيد عن أنفسهم باختيارهم نسبة كبيرة من الفقرات المرغوبة اجتماعيًّا.

وعلى الرغم من أن بعض استبيانات الميول كما رأينا تستخدم فقرات جبرية الاختيار أو تشتمل على فقرات تساعد فى التحقق من صدق الاستجابات، إلا أنه لا توجد أدلة بحثية تؤكد فاعلية ذلك فى التغلب على مشكلات الاتجاهات العقلية فى الاستجابات؛ لذلك لا يوصى خبراء القياس باستخدام استبيانات الميول فى أغراض الانتقاء أو على الأقل عدم الاقتصار عليها كأداة قياس وحيدة.

وإذا استُخدمت فى أغراض الانتقاء أو التسكين ينبغى مراعاة ما سبق أن ذكرناه فى الإرشاد والتوجيه، إذ ينبغى توقع أن يكون هناك اختلاف بين أنماط الميول الأربعة التى أشرنا إليها من قبل، وهى الميول التى يُعبِّر عنها الفرد، والميول التى تقيسها الاستبيانات، والميول التى يُستدل عليها من الاختبارات الأخرى، والميول الفعلية التى تنعكس فى سلوك الفرد. فالفرد يمكن أن يؤدى أنشطة معينة كهواية، ولكنه لا يميل إليها كمهنة. كما أن المصطلحات المتعلقة بالمهن أو تسمياتها، أو مواصفاتها تختلف باختلاف الثقافات بل وباختلاف الاستبيانات، فالمقياس الفرعى المتعلق بمهنة المزارع فى الطبعة الأولى فى مقياس سترونج استُمدت معاييره من أفراد تخرَّج معظمهم من كلية الزراعة فى ولايتين مختلفتين، بينما استُمدت معايير المقياس نفسه فى استبيان كيودر من أفراد تخرج نصفهم فقط فى كلية الزراعة من ولايات متعددة.

أى أن كلا منهما استند إلى عينة منتقاة من خريجى كلية الزراعة، وبخاصة استبيان سترونج. كما أن مواصفات هذه المهنة وغيرها تختلف من ثقافة إلى أخرى. لذلك واجه خبراء القياس فى انجلترا مشكلات متعددة عند استخدام هذين الاستبيانين، نظراً لاختلاف المهن ومواصفاتها عن الولايات المتحدة الأمريكية إضافة إلى المشكلات الأخرى، وللتغلب على ذلك أُعدَّ فى إنجلترا عدة مقاييس بديلة، وفى استراليا أُعدَّ مقياس واحد أطلق عليه مقياس روثويل - ميلر Rothwell - Miller Interest Test حيث اشتمل على ١٢ مجموعة من المهن يقوم الفرد بترتيب ميله إليها. كما أن اختيار مهنة معينة يتأثر بتفاعل كثير من المتغيرات والظروف مثل: نوع الفرد وأدواره، وضغوط الأسرة والأصدقاء، والدافعية، وكلفة التدرب على المهنة، والدخل، والقيم الشخصية.

أغراض تربوية: يمكن الإفادة من نتائج استبيانات الميول، وبخاصة المقاييس البسيطة التى يعدها المعلم لطلابه كما سنوضح بعد قليل، فى أغراض تربوية تشخيصية وعلاجية، كما يمكن الإفادة منها كمؤشر فى تقويم فاعلية البرامج التعليمية، فإذا بيَّنت نتائج تطبيق استبيان الميول فى مستهل العام الدراسى أن بعض الطلبة لا يميلون إلى مجال دراسى معين أو ينفرون منه، فإن هذا يمكن أن يكون مؤشراً لبعض الصعوبات التى ربما تكون سببًا فى ذلك. وعندئذ يستطيع المعلم أن يعاون هؤلاء الطلبة فى التغلب على هذه الصعوبات بتقديم أنشطة محببة لهم فى المجال الدراسى الذى لا يميلون إليه.

فمثلاً إذا كان الطالب ينفر من الرياضيات، ولكنه يميل بدرجة كبيرة إلى الدراسات الاجتماعية، فيمكن أن يُقدم المعلم أنشطة اجتماعية تتطلب قدراً من

العمليات الرياضية، ويشجع الطالب على القراءة فيها لكى يجرى النشاط المطلوب. كما يمكن توظيف ميول الطالب الأكاديمية فى أنشطة عملية تتعلق بمهن مرتبطة بميوله فى إطار المحتوى الدراسى.

كذلك يمكن فى تقويم البرامج التعليمية تطبيق استبيانات ميول مناسبة للمستويات العمرية للطلبة إلى جانب اختبارات الجوانب المعرفية، وذلك فى بداية العام الدراسى وقرب منتصفه أو فى نهايته. فإذا حدث تحسن ملحوظ فى ميول الطلبة، وارتفع مستوى تحصيلهم، فإن هذا ربما يكون مؤشراً على فاعلية البرنامج التعليمى، وهذا يؤكد أهمية الجوانب المعرفية والجوانب غير المعرفية، وتفاعلها وانعكاس أثرها فى أداء الطلبة وسلوكياتهم.

تقويم ميول الطلبة فى الصف الدراسى:

إن استبيانات الميول المقننة التى أوضحناها فيما سبق تفيد فى التنبؤ بالنجاح فى مجال مهنى أو عمل معين والرضا عن العمل وذلك لتوافر معايير لهذه الاستبيانات يمكن فى ضوئها تفسير الدرجات، غير أن المعلم ربما يجد من الأفضل أن يقوم ببناء استبيانات لقياس ميول طلابه إذا لم تكن الاستبيانات المقننة متوافرة وغير ملائمة للطلاب. ولعل الجهد الذى يبذله المعلم فى بناء استبيانات الميول يمكن الإفادة منه فى تقويم الطلبة والتحقق من فاعلية البرنامج التعليمى فى التأثير فى الجوانب الوجدانية لديهم.

فقياس ميول الطلبة يساعد المعلم فى تعرُّف هذه الميول وحفز الميول المرغوبة وتنميتها، وتقديم مثيرات تساعد فى اكتساب ميول جديدة، ومحاولة استبعاد الميول غير المرغوبة، وجعل مجالات الميول جزءاً متكاملاً من المحتوى الدراسى، وإرشاد الطلبة تربويًا ومهنيًا.

لذلك فإن المعلم ينبغى أن يهتم بتقويم ميول طلابه، وأن يدربهم على كيفية إجراء ذلك بأنفسهم. وأن يضع خطة منظمة لإجراء مسح لميولهم فى مستهل العام الدراسى لكى يتعرَّف توجه الميول بعامة. وجمع أدلة تتعلق بالعمليات التى تحدث عند محاولة تعزيز الميول المرغوبة والتخلص من الميول غير المرغوبة، وتعرُّف التغيرات التى تحدث فى أنماط الميول لدى الطلبة، وإعادة مسح الميول فى نهاية العام الدراسى لتحديد التغيرات المرغوبة التى حدثت، وجمع معلومات عن ميول كل طالب للإفادة منها فى التوجيه التربوى والمهنى.

ولكى يحقق المعلم ذلك عليه أن يجمع بيانات دقيقة وموضوعية تتعلق بميول الطلاب للإفادة منها فى أغراض التقويم. وعلى الرغم من أنه يستطيع استخدام بعض استبيانات الميول المقننة المتوافرة مثل التى أوضحناها، إلا أنه ينبغى أن يعتمد بدرجة كبيرة على أدوات القياس التى يُعدها بنفسه لطلابه أو بالمشاركة مع زملائه المهتمين بذلك.

أدوات قياس وتقويم ميول الطلبة فى الصف الدراسى:

تتعدد أدوات قياس وتقويم الميول التى يمكن أن يقوم المعلم بتصميمها وبنائها، واستخدامها فى الصفوف المدرسية، ويمكن أن تكون هذه الأدوات بسيطة مثل مناقشة الطلبة فى الفصل حول اهتماماتهم وتفضيلاتهم، والمقابلات الشخصية، وملاحظة سلوكهم، ويمكن أن تكون أكثر تنظيمًا ودقة، مثل قوائم المراجعة، والاستبيانات، وترتيب الأنشطة، وموازين التقدير، والموازنات الثنائية، ومؤشرات الأداء، واختبارات المواقف وغيرها من أدوات القياس. وسوف نوضح فيما يلى بعض هذه الأدوات:

قوائم المراجعة: يمكن استخدام هذه القوائم Checklists فى تسجيل ميول الطالب وتفضيلاته، وكثير من المعلومات المتعلقة بأنشطة معينة؛ لذلك ينبغى اختيار مجالات الميول المراد قياسها وما تتضمنه من أنشطة. ويمكن أن تكون هذه المجالات عريضة بحيث تشتمل على أنشطة من الجوانب المختلفة للمنهج الدراسى، أو من جوانب مقرر دراسى معين.

ويدوّن اسم الطالب فى أعلى القائمة ، والسلوك أو النشاط المطلوب إلى الجانب الأيمن من القائمة، وتُترك أماكن خالية إلى الجانب الأيسر لكى يتسنى للمعلم وضع علامات أمام كل من هذه الأنشطة بما يراه مناسبًا للطالب. ويمكن أن يكون السلوك لفظيًا مثل اشتراك الطالب فى المناقشة أثناء الدرس، أو غير لفظى مثل اشتراكه فى نشاطات الجمعيات المدرسية ، أو الأنشطة الرياضية.

موازين التقدير: تختلف موازين التقدير Rating Scales عن الاختبارات التى تقيس الجوانب المعرفية فى عدم وجود إجابة صحيحة أو خطأ لكل فقرة من الفقرات، ويحدد الطالب ما يحبه أو يفضله من أنشطة، وما لا يحبه أو لا يفضله منها، أو يحدد درجة تفضيله لكل منها على ميزان تقدير متدرج يشتمل على خمس نقاط مثلا كالتالى:

	أحب جداً	أحب	غير متأكد	لا أحب	لا أحب مطلقاً
قراءة كتب علمية	☐	☐	☐	☐	☐
حضور محاضرات علمية	☐	☐	☐	☐	☐
مشاهدة أفلام علمية	☐	☐	☐	☐	☐
إجراء تجارب علمية	☐	☐	☐	☐	☐

وتقدَّر درجة لكل تفضيل مقابل لكل فقرة، ومتوسط مجموع الدرجات يدل على التفضيل العام للطالب لمجال ميل معين، وهذا يتطلب أن تكون جميع الفقرات متعلقة بهذا المجال. أما إذا كانت الفقرات تقيس مجالات ميول مختلفة، فلا يجوز إيجاد المتوسط العام لجميع الفقرات، وإنما يمكن إيجاد متوسط تقديرات جميع طلاب فصل ما في كل فقرة من الفقرات التي يشتمل عليها ميزان التقدير. ويمكن استخدام أساليب أكثر دقة في بناء موازين التقدير سوف نوضحها في الفصل القادم.

المقابلات الشخصية: يستطيع المعلم إجراء المقابلات الشخصية Interviews في الحصول على معلومات تتصف بالسرية والثقة بينه وبين الطالب ويمكن أن تُعد خطة المقابلات الشخصية مسبقًا أو أن تكون غير سابقة الإعداد، وعادة تجرى المقابلة الشخصية وجهًا لوجه بين المعلم والطالب، وبذلك تفيد في الحصول على معلومات عن الطالب تساعد المعلم في إرشاد الطالب وتوجيهه. ولكن يجب التأكد من أن الطالب لا يجيب عن الأسئلة التي يوجهها إليه المعلم بما يعتقد أنه يحاول إرضاءه، والمقابلات الشخصية تناسب بدرجة أكبر التلاميذ الصغار الذين لم يتمكنوا بعد من مهارات القراءة والكتابة.

ملاحظة السلوك: يمكن أن يستعين المعلم بطرق الملاحظة Observation للحصول على معلومات صادقة عن ميول الطالب وتفضيلاته لأنشطة معينة. لذلك يجب أن تكون هذه الملاحظات واقعية وكمية قدر الإمكان لكي يستطيع المعلم الإفادة من نتائجها. ويجدر ملاحظة السلوك التلقائي للطالب في مواقف طبيعية لكي يستجيب الطالب استجابة تعبر عن ميوله بالفعل. كما يستطيع المعلم أن يستخدم مؤشرات الأداء Performance Indicators في المواقف التي يود أن يُظهر فيها الطالب ميله من خلال ما يختاره ويؤديه من أنشطة حرة أو محددة.

ترتيب الأنشطة: يمكن أن يُعد المعلم قائمة بأنشطة متنوعة ويطلب من الطالب ترتيبها Ranking بحسب درجة ميله لكل منها. وينبغي أن يكون عدد الأنشطة التي تتضمنها القائمة مناسبًا للمستوى العمري للطالب، فالتلميذ الصغير يستطيع ترتيب عدد

محدود من الأنشطة، بينما يستطيع الطالب في المرحلة الثانوية مثلاً ترتيب عدد أكبر من الأنشطة المتجانسة نسبيًا. ويمكن أن يُعيِّن الطالب الرقم(١) للنشاط الذى يُفضله بدرجة كبيرة، والرقم (٢) للنشاط الذى يُفضله بدرجة أقل، وهكذا.

ويمكن أن يوجد المعلم متوسط رتب الفصل فى كل نشاط متضمن فى القائمة، وبذلك يستطيع تقويم ميول طلبة الفصل بعامة وتفضيلاتهم، وينبغى مراعاة أن هذه الأرقام تدل على التفضيل النسبى لكل طالب.

الاستجابات الحرة: يمكن أن يعبِّر الطالب عن ميوله من خلال استجاباته الحرة Free - Responses لمجموعة من الأسئلة التى تتطلب منه كتابة مقال مثلاً عن هواياته أو الأنشطة التى يفضل ممارستها، أو يمكن تحديد مجال الأسئلة قليلاً بحيث يكمل الطالب جملة ناقصة أو سؤالا يتطلب إجابة قصيرة مثل:

أفضِّل قضاء وقت فراغى فى ....

أميل إلى مادة ....

أحب الأنشطة التى تتطلب ....

ما الأنشطة التى تنال اهتمامك غالبًا؟ ....

ويتطلب بناء معظم الأدوات السابقة تحديد مجالات الميول أولا، وتحليلها إلى مكوناتها السلوكية بحيث يمكن قياسها وتسجيل نتائجها بحيث يُستدل منها على إيجابية الميل أو سلبيته أو درجته. فمثلاً إذا كان المعلم مهتمًا بتقويم ميل الطالب نحو دراسة العلوم مثلاً، فإنه يمكنه تحليل هذا الميل إلى مجموعة من الأنشطة التى يمارسها الطالب الذى لديه الميل، مثل:

(١) يحرص على حضور ومتابعة درس العلوم.

(٢) ينتقى الكتب التى تعالج موضوعات علمية عند زيارته لمكتبة المدرسة.

(٣) يقضى وقت فراغه فى قراءة المجلات التى تتناول موضوعات علمية.

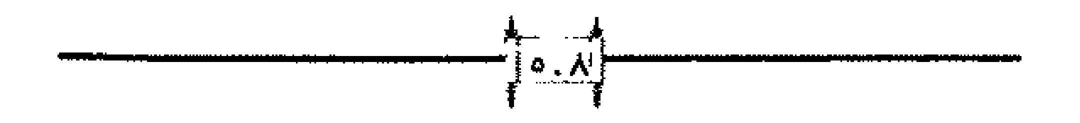
(٤) يحرص على المشاركة فى الرحلات العلمية التى تنظِّمها المدرسة.

(٥) يمارس هوايات علمية.

(٦) يشترك فى الجمعية العلمية بالمدرسة.

(٧) يشارك بانتظام فى أنشطة الجمعية العلمية.

ويمكن أن يسجل المعلم ملاحظاته عن كل طالب فى قوائم أو بطاقات ملاحظة تُصمَّم لهذا الغرض.



وفى ختـام هذا الجزء ينبغى الإشـارة إلى أن مقاييس المـيول تُستخـدم أيضا فى البرامج التربوية التى تهدف لتنمية الوعى المهنى لدى الطلبة فى المرحلة الثانوية، وربما قبل ذلك. وتهتم الدول المتطورة بتصميم برامج منظمة للتوجيه التربوى والمهنى تستند إلى نظم الحـاسوب. وتُعد مقـاييس الميول أحـد مكونات هذه البرامـج، وتهدف هذه البرامج لتنمية مهارات الطلبة فى اتخاذ القرارات المهنية، والنضج المهنى، والتأكد من مناسبة الاختيار المهنى.

ويُعد ذلـك من المجالات البـحثية الـنامية حيـث تتناول مجالات قـياس إدراك الطالب لحاجـاته وعلاقته بعمـلية اتخاذ القرارات المـهنية، ومدى تأكده من اخـتياراته المهنية، والمشاعر المتعلقة بكفاية الطالب وفاعليته، وثقته فى المهام والسلوك المرتبط باتخاذ القرارات المهنية، وتصميم أدوات قياس لهذه المجالات.

خلاصة

تُعد الميول من الجوانب غير المعرفية للشخصية، والتعرف على ميول الأفراد يُعد من الأمور المهمة التي تُوجَّه في ضوئها البرامج التعليمية والتدريبية لتنمية الميول المرغوبة، وبخاصة الميول المهنية. وتتنوع أنماط الميول وأساليب قياسها وتقويمها. فالميول يمكن أن يعبِّر عنها الفرد شفاهة أثناء المقابلة الشخصية أو خلال مواقف الحوار، ويمكن أن تنعكس في سلوكه بحيث يمكن ملاحظة السلوك وتسجيله، ويمكن أن يُستدل عليها من اختبارات التحصيل بافتراض أن التحصيل المرتفع في مجال دراسي معين يرتبط بميل الطالب نحو هذا المجال، وإن كانت الدراسات لم تؤكد صحة هذا الافتراض، نظرًا لتأثير التحصيل بمتغيرات أخرى متعددة. كما تقاس الميول باستخدام الاستبيانات المقننة التي يمكن بناؤها وفق أساليب متعددة تستند إلى أسس نظرية أو منطقية، أو بناء فقرات متجانسة، أو إلى التحقق الإمبيريقي. وقد قدَّمنا أمثلة متعددة لاستبيانات مقننة استند بناؤها إلى كل من هذه الأساليب، ومن أهمها استبيان سترونج - كامبل، واستبيانات كيودر وغيرها.

وعلى الرغم من المشكلات التي تتعلق بهذه الاستبيانات، وبخاصة مشكلات الاتجاهات العقلية في الاستجابات مثل التزييف، والإذعان أو الموافقة، ومراعاة القبول الاجتماعي، والاستجابات الشاذة، ومشكلة تحيز الاستبيانات للذكور، وغير ذلك، فإن هذه الاستبيانات المقننة تُعد أدوات قياس أساسية في برامج الإرشاد والتوجيه التربوي والمهني،ويُستفاد منها إلى حد ما في برامج الانتقاء والتسكين.

وتهتم كثير من الاستبيانات المقننة بقياس الميول المهنية بدرجة أساسية، وبخاصة المجالات التخصصية، غير أن بعض الاستبيانات اهتمت بقياس الميول في مجالات العمل التي تتطلب مهارة أو شبه مهارة مثل الحرف المختلفة. ونظرًا لأهمية تعرُّف ميول الطلبة لتزويدهم بصورة واضحة عن مجالات اهتماماتهم الأكاديمية وتوجهاتهم المهنية، فإن المعلم يستطيع تصميم وبناء أنواع متعددة من مقاييس الميول المناسبة لطلابه مثل قوائم المراجعة، وموازين التقدير، والمقابلات الشخصية، وملاحظة السلوك، وترتيب الأنشطة،والاستجابات الحرة، كما أن قياس الميول يمكن أن يمثِّل جزءًا أساسيًّا من تقويم البرامج التعليمية لأسباب متعددة. إذ يمكن اعتبار

الميول متغيرًا تابعًا، وذلك عندما نود معرفة ما إذا كان للبرنامج التعليمي تأثير في ميول الطلبة، كما يمكن اعتبارها متغيرا تفاعليًا، وذلك عندما نود معرفة ما إذا كان البرنامج فاعلاً بالنسبة للطلبة من ذوى الميول المرتفعة في مجال معين عن أقرانهم من ذوى الميول في مجالات أخرى،لذلك ينبغي قياس الميول قبل بدء البرنامج التعليمي.

كذلك ربما نود قياس الميول لمعرفة تأثيراتها الجانبية، فالبرنامج التعليمي لا يهدف فقط لتنمية الكفايات المعرفية لدى الطلبة ، وإنما أيضًا زيادة اهتمامهم وميلهم نحو مجال دراسي معين وما يرتبط به من المهن.

وعمومًا فإن أي أداة قياس منفردة لا تكون كافية لتقييم الميول باعتبارها ظاهرة مركبة. فكل أداة قياس اعتُني ببنائها وتقنينها تضيف معلومات مسهمة يمكن الاسترشاد بها في عمليات التقييم والإرشاد والتوجيه التربوى والمهني، فهذه المعلومات تزداد فائدتها إذا روُعي أيضًا خلفية الفرد ودافعيته وظروفه الحالية.

الفصل الثانى عشر

قياس الاتجاهات

* مقدمة
* قياس الاتجاه من منظور تاريخى
* مفهوم الاتجاه وأهميته
* علاقة مفهوم الاتجاهات بغيره من المفاهيم الوجدانية
* مكونات الاتجاهات
* خصائص الاتجاهات
* الأسس المنطقية لقياس الاتجاهات
* تصنيف مقاييس الاتجاهات
* أساليب بناء موازين الاتجاهات
* أسلوب الفترات المتساوية ظاهريًا (ثيرستون)
* أسلوب التقدير الجمعى (ليكرت)
* الأسلوب التراكمى (جتمان)
* بعض الأساليب الخاصة
* أسلوب تمايز معانى المفاهيم
* أسلوب التصنيف الترتيبى
* أسلوب البَسْط
* أساليب قياس العلاقات الاجتماعية
* أساليب بناء موازين الاتجاهات متعددة الأبعاد
* تعقيب عام على أساليب بناء موازين الاتجاهات
* إرشادات عامة لبناء موازين الاتجاهات
* أنواع أخرى من مقاييس الاتجاهات
* بعض مشكلات قياس الاتجاهات
* بعض استخدامات مقاييس الاتجاهات

مقدمة :

أوضحنا في الفصل السابق أن الجوانب غير المعرفية أو الوجدانية لشخصية الفرد لا تقل أهمية عن الجوانب المعرفية؛ نظرًا لأنها ترتبط بحاجاته ورغباته ومشاعره ودوافعه النفسية، وتُسهم في نجاحه الأكاديمي والمهني، وتناولنا مفهوم الميول الذي يُمثل أحد الجوانب الوجدانية للسلوك الإنساني، حيث يشير إلى تفضيلات الفرد لأنشطة معينة ينعكس أثرها في سلوكه، وناقشنا أساليب ومنهجيات قياس الميول وما يتعلق بها من مشكلات، وكيفيه الاسترشاد بنتائج القياس في مجالات الإرشاد والتوجيه التربوي والمهني، وفي الأغراض التربوية المختلفة. وأحد الجوانب الوجدانية الأخرى المهمة التي توجه سلوك الفرد وعلاقاته الشخصية والاجتماعية تتعلق بمفهوم الاتجاهات Attitudes .

فالاتجاهات تُعد بمثابة مؤشرات نتوقع في ضوئها سلوكًا معينًا مميزًا للفرد في مواقف لاحقة، فاتجاه الطلاب نحو الكتب المدرسية ربما يؤثر في قدرتهم على تعلم القراءة، واتجاههم نحو المدرسة وبرامجها ربما يؤثر في سلوكهم وقدرتهم على التعلم في المدرسة، واتجاه الفرد نحو المؤسسة التي يعمل بها، والمشرفين على العمل، والآلات والأجهزة التي يستخدمها يؤثر في جودة أدائه. فكما أن الفرد ربما يميل إلى أنشطة معينة، أي يفضل ممارستها، كذلك ربما يكون لديه اتجاه إيجابي أو سلبي بدرجة ما نحو مؤسسات أو مجموعات من الأفراد، أو شخصيات، أو موضوعات، أو ممارسات، أو عادات، أو مفاهيم، أو أفكار معينة.

لذلك يمكن التمييز بين مفهومي الميول والاتجاهات من حيث الإطار المرجعي الذي يشير إليه كل منهما على الرغم من أنهما ليسا مستقلين استقلالا تامًا. إذ ربما يكون الفرد لديه ميل لتنسيق الزهور (نشاط) واتجاه إيجابي نحو الزهور بعامة (أشياء)، وعلى العكس من ذلك ربما يكون لديه ميل للتدريس (نشاط) واتجاه سلبي نحو الأطفال (جماعة من الأفراد). فالاتجاه Attitude كما يعرّفه البورت (Allport, 1954) يُعد استعدادًا عصبيًا نفسيًا تُنظمه الخبرة ويؤثر في استجابات الفرد لجميع المواقف والموضوعات المتعلقة به، ولكن نظرًا لأنه لا يوجد اتجاه واحد يشمل جميع المواقف أو الموضوعات أو الجماعات أو الأشخاص، بل إن كل متغير معين ربما يؤدي إلى انفعالات خاصة به لدى الفرد، وربما تتباين شدة إيجابية أو سلبية هذه الانفعالات، فإن السلوك الوجداني الذي يعكس اتجاهًا معينًا لدى الفرد يتحدد جزئيًا بالشروط المتوافرة وقت ظهور المثير، وفي إطار الظروف التي ظهر فيها. وسوف نفرد هذا الفصل لمناقشة الاتجاهات وطبيعتها وأبعادها، وعلاقتها بغيرها من السمات أو الخصائص الوجدانية،



والأساليب والطرق المختلفة لقياسها، والمشكلات المتعلقة بذلك، واستخدامات أدواتها في مجالات تقويم الأفراد والبرامج التربوية. وسوف نستهل هذا الفصل بتوضيح التطور التاريخي لقياس الاتجاهات.

قياس الاتجاهات من منظور تاريخي:

إن تطور قياس الاتجاهات لم يحدث بمعزل عن علم النفس الاجتماعي، فحتى أواخر العقد الثاني من القرن العشرين لم يكن هناك سوى محاولات متفرقة للقياس في هذا العلم. فعلى سبيل المثال، أعدّ بوجاردوس Bogardus عام ١٩٢٥ مقياسًا للمسافة الاجتماعية Social Distance Scale لتقييم درجات التقبل للأفراد من جنسيات مختلفة في الولايات المتحدة الأمريكية، ويتطلب من المستجيب أن يبين ما إذا كان يتقبل هؤلاء الأفراد كمواطنين عاديين، أو كجيران، أو كأصدقاء، أو كأزواج. وفي عام ١٩٢٥ قام ألبورت وهارتمان Allport & Hartman بدراسة لتعرُّف اتجاهات الطلاب نحو مجالات سياسية مختلفة، وأوضحا الحاجة إلى القياس المقنن للاتجاهات. لذلك أعدّا مجموعة من الفقرات المتعلقة بكل من هذه المجالات، وطلبا من ستة محكمين تقدير رتب لهذه الفقرات، وأوجدا متوسط رتب كل فقرة لتمثيل القيم المناظرة لكل منها، وتطلبت استجابة الطلاب اختيار الفقرة التي تعبّر عن رأيهم بدرجة أفضل حيث قُدِّمت لهم الفقرات مرتبة وفقًا لدرجة أفضليتها، وقد لفتت هذه الدراسة نظر ثيرستون L. Thurstone الذي كان يُعد في ذلك الوقت أسلوب الحكم المقارن Comparative Judgment وغيره من الأساليب المتعلقة بالتحليل السيكوفيزيائي الذي استند إليه في اشتقاق موازين سيكولوجية تتعلق بخصائص المثيرات التي لا يناظرها خصائص فيزيائية يمكن قياسها مثل الخصائص الحسية والجمالية. وقد طوّر هذا الأسلوب وطبقه على بعض فقرات مقياس ألبورت وهارتمان، وبذلك توصّل إلى أساس منطقي لتوزيع آراء الطلاب في المقياس بحيث تكون الفترات المتساوية على المقياس تمثل فروقًا ظاهرية متساوية في الآراء.

وقد أدرك ثيرستون أهمية هذا الأسلوب في قياس الاتجاهات الاجتماعية، ووضح ذلك في مقالته الرائدة بعنوان «الاتجاهات يمكن قياسها» التي نُشرت عام ١٩٢٨، حيث تناول فيها الأساس المنطقي لأسلوبه الجديد الذي أطلق عليه «الفترات المتساوية ظاهريًا Equal Appearing Intervals». ويتطلب هذا الأسلوب أن يقوم المستجيب بتصنيف الفقرات في مجموعة من الأقسام المرتبة التي يبلغ عددها عادة ٩ أو ١١ قسمًا، بحيث إن الأقسام تبدو أنها تمثل تدريجًا للآراء متساوي الفترات. ومن ثم يُستخدم أسلوب الحكم المقارن أو أي أسلوب رياضي مماثل للتوصل إلى قيم الميزان المتدرج Scale Values تُمثل الأفضلية النسبية لكل فقرة تجاه الموضوع أو

القضية التى تتضمنها الفقرة. وقد أسهم هذا الأسلوب فى تطوير البحث فى مجال الاتجاهات الاجتماعية Social Attitudes، كما أسهم فى بناء العديد من مقاييس الاتجاهات فى مجالات متعددة، وطُبقت هذه المقاييس على جماعات متنوعة، ومن أهمها مقاييس ريمرز ومعاونيه Remmers et al. للاتجاهات نحو المجالات الدراسية، والمؤسسات. وقد اقترح ليكرت R. Likert عام ١٩٣٢ مدخلا مختلفًا لقياس الاتجاهات والجماعات العرقية والأنشطة المتعلقة بالأسرة، والممارسات الاجتماعية والأخلاقية، والمهن المختلفة، والإعلانات، حيث بدأ إعدادها عام ١٩٣٠. واستند فى ذلك إلى أساليب تحليل المفردات المستخدمة فى مجالات الاختبارات العقلية. وتطبق هذه الأساليب مباشرة على استجابات الأفراد المتعلقة بالموافقة أو عدم الموافقة على كل فقرة من فقرات مقاييس الاتجاهات دون الحاجة إلى استخدام أحكام الخبراء فى الحكم على درجة أفضلية الفقرات. ويُطلق على هذا الأسلوب «التقديرات الجمعية Summated Ratings»، حيث يُطلب من المستجيب أن يُحدد درجة موافقته أو عدم موافقته لكل فقرة من فقرات المقياس على ميزان متدرج يتكون من خمسة أرقام لكل منها وزن يبدأ من نقطة واحدة إلى خمس نقاط، مثل : أوافق بشدة، أوافق، غير متأكد، لا أوافق، لا أوافق على الإطلاق. وتُجمع النقاط المناظرة لاستجابات الفرد على الفقرات للحصول على درجته فى المقياس.

وفى عام ١٩٤٤ قدم جتمان L. Guttman أسلوبًا جديدًا وضّحه فى مقال بعنوان «أسس تدريج البيانات الكيفية» ويتعلق بتحليل مصفوفات بيانات الموازين المدرجة Scalogram Analysis ، حيث يتم ترتيب المفردات والأفراد وفقًا لاستجاباتهم، فالفرد الذى يوافق على فقرة معينة يوافق أيضًا على جميع الفقرات الأدنى منها فى الرتبة. وتلا ذلك الأسلوب الذى اقترحه لازارسفلد Lazarsfeld عام ١٩٥٠ المتعلق بتحليل البنية الكامنة للمتغيرات Latent Structure Analysis. ويُعد هذا الأسلوب نموذجًا رياضيًّا عامًّا يربط احتمال الاستجابة لفقرة معينة بموقعها على متغير كامن ينطوى عليها مثل الاتجاه Attitude. والحقيقة أن أساليب ليكرت، وجتمان، والتحليل العاملى يمكن أن تعد حالات خاصة من نموذج لازارسفلد إذا تحققت شروط معينة فى توزيع متغير الاتجاه، وشكل الدالة التى تربط احتمال الاستجابة بالمتغير الكامن (الاتجاه)، ومع هذا فإن أسلوب لازرسفلد لا يُستخدم إلا فى الأغراض البحثية فى مجال الاتجاهات. كما قدم كومبس Coombs عام ١٩٥٠ نظرية البيانات Theory of Data التى تُعد نظرية عامة لتدريج استجابات التفضيل Preferences Scaling، وفى عام ١٩٥٠ أيضًا نُشرت دراسة أدورنو ومعاونيه Adorno et al. المتعلقة بالشخصية السلطوية The Authoritarian Personality تناولت منهجية جديدة تضم استخدام أسئلة حقائق، ومقاييس تجمع الرأى والاتجاه فى أسئلة إسقاطية مفتوحة النهاية،

ومقابلات كلينيكية، حيث يرون أن الأساليب المتعددة غير المباشرة Disguised مهمة للتمييز الصادق بين اتجاهات الأفراد.

مفهوم الإتجاه وأهميته:

نال مفهوم «الاتجاه Attitude» اهتمام علماء النفس الاجتماعي وعلماء القياس نظرًا لتأثر سلوك الأفراد تأثرًا ملحوظًا بالاتجاهات، مما يؤثر بدوره في العلاقات الإنسانية التفاعلية بين أفراد المجتمع الواحد والمجتمعات المختلفة. ويعتمد استقرار هذه العلاقات على مدى تأثير أنماط الاتجاهات السائدة في المجتمع في حياة الأفراد دون ضغوط أو توترات. لذلك اهتم خبراء التربية أيضًا باتجاهات الطلاب والتحقق من فاعلية العملية التربوية في تكوين أو تنمية أنماط اتجاهات تُسهم في إحداث تفاعل إيجابي وتكامل بين القوى المختلفة في المجتمع. فإذا كان المعلم يعمل على تنمية وإثراء تفكير طلابه، فإن عليه أن يدرك أن التفكير لا يحدث في فراغ، وإنما يُوظف أنماط اتجاهات الطالب وخبرته، فليس المهم فقط كيف يفكر الطالب، وإنما المهم هو فيم يفكر. كما أن التكيف الاجتماعي الانفعالي للطالب يتأثر بتقييمه لسلوكه ورغباته ومشاعره في علاقتها بالقيم الاجتماعية كما يدركها بنفسه. فاتجاهه نحو البيئة ونحو زملائه ومعلميه والمؤسسات المختلفة في مجتمعه يؤثر تأثيرًا أساسيًا في توجهات تفكيره إما بالإيجاب أو بالسلب. لذلك فإن تكامل الجوانب العقلية والوجدانية للطلاب يتعلق باتجاهاتهم بقدر تعلقه بالتكامل الاجتماعي وتماسك المجتمع. ولهذا فإن مفهوم الاتجاه يُعد من المفاهيم السيكولوجية الاجتماعية التربوية المهمة، حيث إنه يُعد من المكونات الوجدانية الأساسية.

وقد تباينت تعريفات الاتجاه تباينًا ملحوظًا، فقد ذكر ألبورت Allport منذ أكثر من نصف قرن من الزمان سبعة عشر تعريفًا مختلفًا للاتجاه، كما وجد ماكجواير McGuire في مسحه لتعريفات الاتجاه عام ١٩٦٩ أن هناك ما يقرب من ثلاثين تعريفًا. وحاول ألبورت استخلاص العناصر الأساسية للاتجاه من التعريفات التي ذكرها، وتوصل إلى ثلاثة عناصر هي: التهيؤ أو الاستعداد لاستجابات القبول أو الرفض، والاستجابات التي تُنظمها الخبرة، وتحفز الاستجابات في وجود جميع الأشياء والمواقف التي يتعلق الاتجاه بها.

وعلى الرغم من تعدد التعريفات بمرور الزمن، إلا أن معظمها يجمعها العناصر الأساسية الثلاثة السابقة. ولعل الاهتمام بتعريف الاتجاه تعريفًا إجرائيًا يساعد في تحديدأساليب القياس المناسبة.

وخلاصة التعريفات التى احتوت العناصر الثلاثة السابقة يمكن صياغتها فى أن الاتجاه تكوين افتراضى يتضمن استجابة محفِّزة عندما يواجه الفرد مثيرات اجتماعية بارزة، وتتميز هذه الاستجابة بخصائص تقييمية. ويحتاج هذا التعريف إلى مزيد من التوضيح كما يلى:

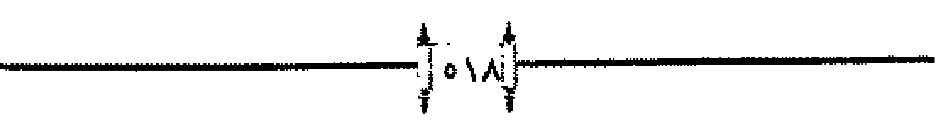
(١) الاتجاه متغير مثل غيره من المتغيرات السيكولوجية، أى أنه متغير أو سمة كامنة أو تكوين فرضي، وليس متغيراً يمكن ملاحظته ملاحظة مباشرة، وبهذا لا يشير مفهوم الاتجاه إلى فعل واحد أو استجابة واحدة للفرد، وإنما يُعد تجريداً مشتقًا من مجموعة كبيرة من الأفعال أو الاستجابات المترابطة، أى نستدل عليه من هذه المجموعة من الاستجابات التى ينبغى أن تكون متسقة. وهذا يعنى أنه إذا كان الأفراد الذين يعبِّرون عن رفضهم لقضية معينة يُحتمل أيضًا أن يعارضوا ما يتعلق بهذه القضية وما يترتب عليها، فإنه يمكننا عندئذ أن نستدل على اتجاههم نحو هذه القضية.

فالاتجاه إذن كمتغير كامن يُستخدم لوصف اتساق أو تغاير عدد من الاستجابات المختلفة لمثيرات تنتمى إلى مجموعة عامة واحدة، أى أن المتغير الكامن (الاتجاه) يتوسط الاستجابات والمثيرات؛ لذلك تتغاير الاستجابات لتعلقها بالمتغير الكامن نفسه. فأهمية هذا المتغير الكامن أو الفرضى ترجع إلى أنه توجد مجموعة من البيانات المتعلقة بالاستجابات الملاحظة. ومثل هذا المتغير يطلق عليه «سمة Trait» أو «متغير وسيط Intervening Variable» أو « عامل Factor »أما الاستجابات الملاحظة فيُطلق عليها متغيرات ظاهرة أو صريحة أو مؤشرات Indicants.

(٢) الاتجاه الذى لدى الفرد يجعله يلاحظ ويستجيب بطريقة انتقائية، فإذا كان الطالب مثلا لديه اتجاه إيجابى نحو معلميه، فإنه ربما يلاحظ مدى معاونتهم لطلابهم، وربما يُفسرر البقاء بعد انتهاء اليوم المدرسى، لمتابعة المناقشة معهم، بينما إذا كان اتجاهه سلبيًا نحوهم، فإنه ربما يلاحظ مدى سيطرتهم على الطلاب، وربما يفتعل الأعذار للتغيب عن المدرسة، لذلك فإن الملاحظة الدقيقة للقرائن التى يلمسها الطالب والأفعال التى يقوم بها بعد ذلك ربما تفيد فى الاستدلال على اتجاهاته.

(٣) الاستجابات التى يتضمنها الاتجاه تسمح بالاستدلال عليه منها عندما يواجه الفرد مثيرات مناسبة، وهذه المثيرات يمكن أن تكون لفظية يعبِّر عنها الفرد فى استجابته على مقاييس تتطلب الورقة والقلم، أو غير لفظية يمكن ملاحظتها من سلوك الفرد فى مواقف فعلية أو محاكاة، وفى جميع الأحوال يُستدل على الاتجاهات من نمط استجابات الفرد لفئة أو مجموعة من المثيرات.

(٤) الاتجاه يُعد تقييمًا شخصيًا لمواقف أو أشخاص أو مؤسسات أو أفكار، حيث يحتوى على عناصر موجبة مثل الموافقة أو الاقتراب أو التفضيل، وعناصر سالبة مثل عدم الموافقة أو محاولة الابتعاد أو الكراهية.



والخلاصة أنه يمكن اعتبار الاتجاه انفعالاً معتدل الشدة يهيئ الفرد أو يجعله مستعداً للاستجابة المتسقة التى تدل على الموافقة (التأييد) أو عدم الموافقة (عدم التأييد) عندما يواجه موضوع الاتجاه. وهذا التعريف يساعد فى تمييز مفهوم الاتجاه عن غيره من المفاهيم المتعلقة به والتى سنوضحها فيما يلى:

علاقة مفهوم الاتجاه بغيره من المفاهيم المتعلقة به:

لكى يتضح مفهوم الاتجاه بدرجة أفضل، ربما يكون من المناسب أن نوضح علاقته بغيره من المفاهيم المتعلقة به. وسوف نقتصر على أهم هذه المفاهيم وبخاصة التى يحدث خلط بينها وبين مفهوم الاتجاه، وهى: الميول، والآراء، والمعتقدات، والقيم، والدوافع، والعادات والسمات، والروح المعنوية.

الميول : تعكس الميول Interests المتمثلة فى السلوك ما يجذب الفرد أو يُبعد اهتمامه عن أنشطة معينة، أى أن الميول تعبِّر عن تفضيلات الفرد للقيام بأنشطة دون سواها، فى حين أن الاتجاهات لا تقتصر على المشاعر، حيث إنها تتضمن أحكاماً قيمية، أى استجابات القبول أو الرفض.

وعلى الرغم من أنه لا ينبغى النظر إلى الميول على أنها منفصلة عن الاتجاهات، إلا أنه ربما يميل فرد بدرجة كبيرة إلى نشاط معين على الرغم من أن حكمه القيمى على ما يتضمنه هذا النشاط يكون سلبياً.

الآراء : هناك أوجه تشابه بين الآراء Opinions والاتجاهات، غير أن مفهوم «الرأى» يشير عادة إلى الاقتراع Opinion Polls ، حيث يُطلب من الفرد التعبير عن مشاعر التفضيل تجاه كل فقرة على حدة، وليس على مجموعة متجانسة من فقرات استبيان أو مقياس معين، مثل ما تقديرك لنظام اليوم الدراسى الكامل؟ (ممتاز، جيد، مناسب، غير مناسب). والآراء تتضمن مشاعر أقل مما تتضمنه الاتجاهات، وأقل منها موضوعية، وتعتمد بدرجة أكبر على حقائق أو معلومات لدى الفرد فيما يتعلق بالموضوع الذى يُقترع عليه، كما أن الآراء لفظية، بينما الاتجاهات تتوسط أحياناً بعمليات غير لفظية أو لا شعورية، أى أن الآراء تنعكس فى استجابات لفظية بينما الاتجاهات تُعد استعدادات سابقة أو تهيؤاً Predispositions.

المُعتقدات : تتعلق بمستوى قبول معين لقضية مرتبطة بخصائص شيء أو حدث معين. ويعرِّف إنجليش وإنجليش (English & English, 1958) المعتقدات Beliefs بأنها القبول الانفعالى لقضية أو مبدأ يعتبره الفرد أساسياً. وتشير معظم التعريفات إلى أن المُعتقد يصبح اتجاهاً إذا صاحبته مشاعر معينة تعكس تقييم تفضيل الفرد لخصائص الشيء أو الحدث المعين، وعندئذ يعبِّر عن الاتجاه بمجموع هذه المعتقدات. غير أن

هناك من يرون أن المعتقدات أقل قابلية للتغير من الآراء بل والاتجاهات، ومن أمثلة ذلك الاعتقاد بأن الأرض كروية، أو الاعتقاد بالتضحية الوطنية وغير ذلك.

القيم : يرى سميث (Smith, 1963) أن القيم Values هي تصورات فكرية شخصية لما هو مرغوب ومتعلق بسلوك انتقائي، مثل الحكم على العمل بما يدره من دخل، أو مراعاة الأخلاق في جميع التصرفات، أو احترام العلم والمتعلمين، أي أن القيم تتعلق بأشياء أو مفاهيم مهمة في الحياة. ويمكن أن تكون القيم موجبة(التسامح الاجتماعي، والتطلع العلمي، والمعايير الخلقية) أو سالبة (الجريمة، والتحيز العرقي، وعدم التحمل، والفقر، والجهل). وتتكون القيم من تجمع ميول الفرد وتذوقاته وتفضيلاته واتجاهاته المتعلقة بظواهر مختلفة، وتتضمن مشاعر وانفعالات أقوى من الاتجاهات. وتحتاج القيم إلى وقت طويل لكي تتكون وتتشكل، ومن الناحية التربوية، فإن القيم تنمو وتتكون وتُتعلم عن طريق ما نقوله ونلاحظه ونفعله في البيئة المحيطة بنا.

الدوافع : تتضمن الاتجاهات الدوافع Motives، فإذا كان لدى الطالب اتجاه إيجابي نحو معلم معين، فإن هذا ربما يكون دافعًا له للاقتداء به. وتؤكد بعض التعريفات أن الاتجاهات استعدادات سابقة أو تهيؤ لدافع الاستثارة، فالدافعية إذن مؤقتة، وأكثر نوعية حيث تستهدف تحقق هدف معين، بينما الاتجاه يداوم تأثيره من موقف إلى آخر، ويُحفز للسلوك.

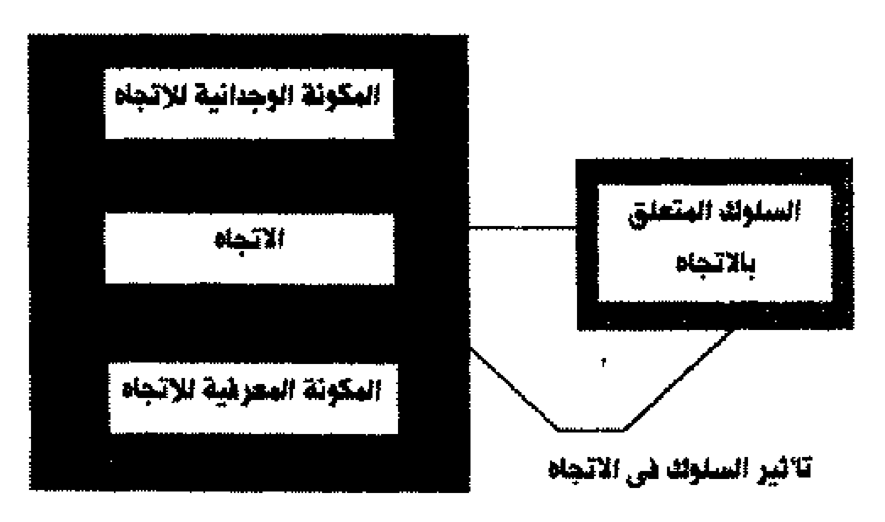
العادات والسمات : تختلف الاتجاهات عن العادات Habits والسمات Traits، فالعادات تشير إلى نزعة الفرد لعمل معين، وتتميز ببنية معقدة مستديمة. والعادات كالاتجاهات تكون مكتسبة، ولكنها لا تتضمن حكمًا تقييميًا على العكس من الاتجاهات، أما السمات فهي استعدادات متسقة، وأقل استقرارًا لاستجابة الفرد بطريقة ما تميزه عن غيره من الأفراد، وتختلف الاتجاهات عن السمات في كونها نوعية التوجه، بينما السمات تتميز بعمومية توجُّه الفرد نحو أشياء متنوعة ومتسعة.

الروح المعنوية : تشير إلى التكامل الانفعالي لجماعة معينة، فالروح المعنوية Morale تُعد نمطًا متكاملاً من الاتجاهات التي تشير إلى أشياء معينة ذات أهمية حيوية للجماعة، ومن أمثلة ذلك الروح المعنوية لطلاب صف معين، أو الروح المعنوية في مؤسسة صناعية، أو عسكرية، أو الروح المعنوية للأفراد في أوقات الحروب. وارتفاع الروح المعنوية يسهم في تحقيق أهداف معينة، وانخفاضها ربما يهدد مصالح حيوية للجماعة.

ومن هذا يتضح تعدد المفاهيم المتعلقة بمفهوم الاتجاه، ويرجع ذلك إلى عدم الاتفاق على معنى ودلالة كثير من المفاهيم الوجدانية، واختلاف الإطار النظرى الذى يستند إليه الباحثون فى هذا المجال، ونواتج التحقق الإمبيريقى الذى يهدف لتحديد أوجه التشابه والاختلاف بين هذه المفاهيم المتعددة. ونظرًا لاختلاف الأسس والمحددات التى يُستند إليها فى تعريف الاتجاه Attitude بمرور الزمن، لذلك ربما لا نتوقع التوصل إلى تعريف نهائى واحد لهذا المفهوم فى الوقت الحاضر. ويُفضل اعتبار جميع هذه المفاهيم أبعادًا متعددة يمكن تعريفها وقياسها بدرجات متفاوتة. ولعل تمييز مفهوم الاتجاه عن غيره من المفاهيم السابقة يعتمد فى وقتنا الحاضر على ما ينال اتفاق الباحثين.

مكونات الاتجاهات:

اتضح لنا مما سبق أن الاتجاهات Attitudes تُعد من خصائص الأفراد، حيث تصف مشاعرهم الموجبة والسالبة تجاه مواقف أو أفكار أو أشخاص أو مؤسسات أو موضوعات معينة. ويرى واجنر (Wagner 1969)، وزيم وإيبسون (Zimmbardo & Ebbeson, 1970) أن الاتجاهات لا تقتصر على المكونة الوجدانية أى المشاعر نحو موضوع ما، وإنما تتضمن أيضًا مكونة معرفية تتعلق بمعارف الفرد أو أفكاره أو إدراكاته أو آرائه أو معتقداته حول الحقائق المرتبطة بموضوع الاتجاه، كما تتضمن مكونة سلوكية تتعلق بأفعال الفرد واستجاباته وسلوكه الملاحظ نحو الموضوع المعين. ويمكن توضيح هذه المكونات الثلاث للاتجاه بالشكل التخطيطى (١٢ـ١) التالى:



شكل (١٢ـ١) يوضح المكونات الوجدانية والمعرفية والسلوكية للاتجاهات وتفاعلها

فمعظم الاتجاهات يتعلمها الأفراد من والديهم ومن المحيطين بهم. ولعل الآراء التى يتم تعزيزها بانتظام من جانب الوالدين والأصدقاء تدعم المكونة المعرفية للاتجاهات Cognitive Component. وبالمثل فإن اقتران المعلومات المتعلقة بموضوعات الاتجاهات بالاستجابات الانفعالية للأفراد ربما يؤدى إلى دعم المكونة الوجدانية Affective Component. كما أن اتجاهات الأفراد تعتمد على اتجاهات أفراد الجماعات الأخرى التى ينتمون إليها، وهذه الجماعات تُسهم فى دعم مكونتى الاتجاه، وتسمى الجماعات المرجعية Reference Groups. وربما تكون كل جماعة مرجعية مناسبة لاتجاهات معينة بالنسبة للفرد، إذ ربما ينسب الفرد اتجاهه نحو موضوع معين، إلى إحدى الجماعات أو المؤسسات المرجعية، بينما ينسب اتجاهًا آخر إلى جماعة أو مؤسسة مرجعية أخرى. والحقيقة أن تحليل الاتجاهات إلى المكونات الثلاث الموضحة بشكل (١٢-١) يشير إلى أن اكتساب الاتجاهات يُعد من العمليات المعقدة. فالاتجاهات تعد بمثابة متغيرات وسيطة Intervening Variables. والسهم المتجه من السلوك إلى الاتجاه كما يوضحه شكل (١٢-١) يُعد بمثابة تغذية راجعة للاتجاه مما يشير إلى حقيقة أن الاتجاهات يمكن أن تتغير بتأثير المكونة السلوكية Behavioral Component.

وكذلك السهم المتجه من المكونة المعرفية يوضح تأثر الاتجاه بها أيضًا، وذلك لأن إدراك الفرد لموضوع الاتجاه ومعلوماته عنه تؤثر فى درجة مشاعره نحوه مما يؤثر بدوره فى سلوكه الملاحظ.

وعلى الرغم من أن هذا النموذج الذى يربط بين مكونات الاتجاه يُعد من النماذج الكلاسيكية، إلا أن الدراسات الإمبيريقية قدمت أدلة تؤيده (Kahn , 1951; Vidulich & Krevanick, 1966 ;Woodrnansee & Cook , 1967) غير أن شو، ورايت (Shaw & Wright,1967) أوضحا أن المكونة المعرفية تُعد الأساس الذى يستند إليه الفرد فى تقييمه لموضوع الاتجاه، وبالتالى للاتجاه ذاته الذى يهيئ الفرد للاستجابة بطريقة معينة. ويرى شو، ورايت أن الاتجاهات يمكن اعتبارها مجموعة من ردود الأفعال الوجدانية نحو موضوع الاتجاه استنادًا إلى مفاهيم ومعلومات ومعتقدات الفرد حول هذا الموضوع، وتجعله أكثر تهيؤًا لكى يسلك بطريقة معينة نحوه. وتُعد المكونة المعرفية الأساس الذى يستند إليه تقييم الفرد لموضوع الاتجاه.

خصائص الاتجاهات:

تتميز الاتجاهات ببعض الخصائص العامة عن غيرها من المفاهيم الوجدانية التى سبق أن أوضحناها. وتفيد هذه الخصائص فى إلقاء مزيد من الضوء على الأبعاد المتعددة لهذا المفهوم. وسوف نوضح فيما يلى أهم هذه الخصائص:

(١) التوجُّه: يُعد من الخصائص الأساسية للاتجاهات، فالتوجُّه Direction يتعلق بما إذا كانت مشاعر أو انفعالات الفرد تجاه موضوع أو قضية معينة موجبة أم سالبة. وهذه الخاصية تُعد من الخصائص المعقدة؛ نظرًا لأن الاتجاهات تتضمن استجابات ظاهرية متنوعة، مثل الثقة، والحب، والتعاون، والموافقة، والقبول، فى مقابل الكراهية، والشك، والانسحاب، والاعتراض، والمهاجمة. ويمكن أن تندرج جميع هذه المشاعر تحت مفهوم عام يتعلق بالمسافة السيكولوجية Psychological Distance بين الفرد وموضوع الاتجاه. فالطالب الذى لديه اتجاه موجب نحو المدرسة يكون موجّهًا توجيهًا إيجابيًّا نحو بعض الأشياء المتعلقة بالمدرسة مثل الطلاب والمعلمين،والجمعيات المدرسية، والمنهج والمواد الدراسية، أما إذا كان لديه اتجاه سالب نحو المدرسة فإنه يحاول الابتعاد عن المدرسة وما يرتبط بها من أنشطة.

غير أن هذه الخاصية تُبرز التساؤل المتعلق بالاتجاهات المحايدة، وما إذا كانت تقع بين الاتجاهات الموجبة والاتجاهات السالبة. فتصورنا للتوجُّه الموجب والسالب على أنهما توجُّهان متضادان، يعنى أن اتجاه الفرد لا يمكن أن يقع على طرفى النقيض فى آن واحد. لذلك ربما يكون من المناسب معالجة درجات التفضيل ودرجات عدم التفضيل على أنهما متمايزان من الناحية النظرية، على الرغم من أنهما بلا شك يكونان مرتبطين من الناحية الإمبيريقية، أى أنه يمكن من الناحية النظرية التصورية اعتبار أن الاتجاه خاصة ثنائية القطب Bipolar Attribute.

(٢) المقدار والشدة: يشير مقدار الاتجاه Magnitude إلى درجة الاتجاه الموجب أو السالب. وعادة يحدث خلط بين مقدار الاتجاه وشدته Intensity ، فشدة الاتجاه تشير إلى أهمية أو قوة مشاعر الفرد تجاه موضوع معين، فبعض الأفراد تكون لديهم مشاعر أو انفعالات أكثر شدة من غيرهم من الأفراد، وكذلك المشاعر ذاتها يمكن أن تتباين فى درجة شدتها، فالكراهية تُعد من المشاعر الأكثر شدة من عدم الحب، ويمكن أن يكون لدى أحد الأفراد اتجاه موجب ولدى فرد آخر اتجاه سالب نحو موضوع ما، ومع هذا تكون شدة اتجاه كل منهما متساوية. وكلما زادت درجة التفضيل أو عدم التفضيل زادت شدة الاتجاه، أما الأفراد المحايدون فإنه يكون لديهم

اتجاه أقل شدة، والحقيقة أن مقدار الاتجاه وشدته يُعدان من الخصائص الأساسية التى تميِّز الاتجاه عن غيره من الجوانب الوجدانية، لذلك تهتم مقاييس الاتجاهات بقياس هاتين الخاصيتين.

(٣) ثنائية المشاعر: تشير إلى مدى تناقض مشاعر الفرد نحو جوانب مختلفة لموضوع الاتجاه نفسه، وتتأتى ثنائية المشاعر Ambivalence من تصور التوجه ثنائى القطب للاتجاه الذى أشرنا إليه، فكلما تساوت مشاعر التفضيل ومشاعر عدم التفضيل ارتفعت درجة تناقض المشاعر.

(٤) المركزية: تشير هذه الخاصة Salience إلى التهيؤ لاستثارة الاتجاه، أى مدى اقتراب الاتجاه من التفكير المباشر للفرد، فمثلاً إذا سألنا بعض الأفراد عن أهم مشكلة تواجه البشرية اليوم، فإنهم ربما يفكرون على الفور فى منع الحروب، بينما إذا سألنا بعضًا آخر ربما يفكر على الفور فى محاربة الإدمان، ويفكر غيرهم مباشرة فى حقوق الأطفال وهكذا، أى أن هذه الخاصة تتعلق بما إذا كان اتجاه معين يعد مركزيًا لتفسير اتجاهات واسعة ومتنوعة.

والاتجاهات المركزية هى تلك التى يهتم بها الفرد، ويكون لديه قدر كبير من المعلومات حول موضوعها. ويمكن قياس هذه الخاصة إذا أتيحت الفرصة للفرد للتعبير الحر عن اتجاهه دون سبر. ونظرًا لأن معظم مقاييس الاتجاهات تشتمل على فقرات تتطلب استجابات محددة مثل «موافق» أو «غير موافق» فإننا لا نستطيع باستخدامها قياس خاصة «المركزية» ، وإنما يمكن قياسها باستخدام أساليب الملاحظة والمقابلات الشخصية حيث يمكن أن يعبر الفرد عن اتجاهاته دون قيد.

(٥) المركزية الوجدانية: تشير إلى الدرجة التى يصبح فيها الفرد انفعاليًا بدرجة كبيرة فى تعبيره عن اتجاهاته نحو موضوع معين. أى أن خاصة المركزية الوجدانية Affective Salience تتعلق بدرجة تقييم الفرد أو انفعاله المتمركز حول موضوع الاتجاه.

(٦) المرونة: تشير إلى مدى قابلية الفرد لتغيير أو تعديل اتجاهاته نحو موضوعات معينة، وعادة ينظر إلى خاصة المرونة Flexibility على أنها سمة عامة للأفراد بدلاً من اعتبارها تشير إلى اتجاه محدد.

(٧) التضمن: فالاتجاه يمكن أن يكون منفصلا عن الجوانب الأخرى، مثل المعتقدات، أو القيم، أو غيره من الاتجاهات، أو يمكن أن يرتبط بهذه الجوانب من خلال الاقتران والتصنيف والتبرير وهكذا. فمفهوم الأم يمكن أن يكون متضمنًا أو مطمورًا Embedded فى شبكة من الترابطات بمفاهيم متمايزة مثل الحماية، ورعاية الطفولة، وغير ذلك.

(٨) الشمول أو المدى : فالطالب يمكن أن يكره بشدة مظهرًا أو مظهرين متعلقين بالمدرسة مثلاً، بينما يمكن أن يكره طالب آخر كل ما يتعلق بالتعليم، أو ربما يكون لدى أحد الأفراد اتجاه سالب نحو الشباب الذى يتميز بخصائص معينة فقط، أو ربما يكون اتجاهه السالب نحو الشباب بعامة الذين تختلف خصائصهم عنه. أى أن الشمول أو المدى Pervasiveness يتعلق بمدى تعميم الاتجاه أو مدى شموله لعناصر موضوع الاتجاه، فإذا علمنا أن فردًا ما متحفظ، فإنه يُمكننا التنبؤ باتجاهاته نحو موضوعات متنوعة واسعة، وكثير من الاتجاهات تتميز بإمكانية تعميمها بدرجة كبيرة.

(٩) التعقُّد المعرفى : يشير إلى تفاصيل المكونة المعرفية للاتجاهات، مثل ثراء المحتوى، أو عدد الأفكار التى لدى الفرد عن موضوع الاتجاه، وأحيانًا يطلق على التعقُّد المعرفى Cognitive Complexity تسميات متعددة مثل: التمايز، والتعدد. وتعتمد درجة التعقد المعرفى إلى حد ما على منظور الفرد عن موضوع الاتجاه، فمنظور الفرد عن الناس بعامة ربما يختلف عن منظوره لفرد معين واحد. فإذا وصف الفرد أحد الأشخاص بأنه حسن الخلق، ومهذب، فإن استخدام هذين الوصفين المترادفين معًا للتعبير عن اتجاه الفرد، ربما يشير إلى اتجاه أكثر تعقدًا مما لو استخدم أحد هذين الوصفين فقط. غير أن هذا التفسير لا يزال موضع خلاف فى بحوث الاتجاهات.

(١٠) الاتساق : تُميز خاصة الاتساق Consistency بين الخصائص الوجدانية ورد الفعل الوجدانى الذى تُحدثه مواقف أو أحداث معينة، فبعض الأفراد تكون استجاباتهم على مقاييس الاتجاهات متسقة اتساقًا ملحوظًا، بينما يستجيب البعض الآخر استجابات متناقضة لموضوع واحد من حيث التفضيل أو عدم التفضيل. لذلك فإنه ينبغى أن تكون هناك درجة ما من اتساق الاستجابات لكى نتمكن من الاستدلال على اتجاه الفرد، وإذا لم تتوافر هذه الدرجة فإن استجابة الفرد ربما تُعزى إلى عوامل خارجة عنه، ولا تُعزى إلى اتجاهه الفعلى. والحقيقة أن خاصة الاتساق تُعد من الخصائص الأساسية للاتجاهات، بالإضافة إلى خاصة مقدار أو شدة الاتجاهات وتوجُّهها، ويمكن ملاحظة ذلك فى تعريف مفهوم الاتجاه الذى سبق أن قدمناه. لذلك تركز موازين الاتجاهات Attitude Scales على قياس هذه الخصائص الثلاثة كما سيتضح لنا فيما بعد.

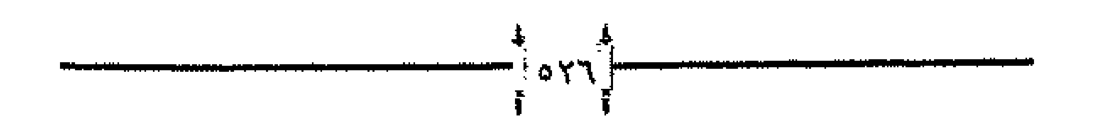
الأسس المنطقية لقياس الاتجاهات :

- سبق أن أوضحنا في الفصل الأول أن القياس النفسي والتربوي يتعلق بجمع ملاحظات أو بيانات عن سلوك الأفراد، وتعيين أرقام تمثل هذه الملاحظات وفقًا لقواعد محددة، وينطبق ذلك على قياس الاتجاهات. وتعتمد إجراءات القياس على الافتراضات النظرية المتعلقة بطبيعة الاتجاه الذى نود قياسه، وعلى العلاقات القائمة بينه وبين السلوك، وكذلك على علاقته بالقواعد المستخدمة في تعيين أرقام للملاحظات السلوكية. فالنماذج المختلفة لبناء الموازين Scaling Models المستخدمة في قياس الاتجاهات تهدف لتحويل مجموعات من الملاحظات أو استجابات الأفراد إلى مؤشرات Indices للاتجاه الذى تنطوى عليه هذه الاستجابات، غير أن تفسير البيانات المستمدة من هذه الموازين يعتمد على العلاقة بين مفهوم الاتجاه الذى يتبناه من يقوم بإعداد الميزان Scale والأسلوب الذى يستخدمه في قياس هذا الاتجاه. ولعل وضوح هذه العلاقة يعد أمرًا ضروريًا لفهم النماذج والأساليب المختلفة لقياس الاتجاهات. وتمر عملية قياس الاتجاهات بعدة مراحل متتالية نوجزها فيما يلي:

(١) تصور مبدئي لطبيعة الاتجاه المراد قياسه، ولعل التعريفات التى قدمناها فيما سبق تُوضح ذلك.

(٢) تحديد الأبعاد المتعلقة بالاتجاه للاستناد إليها في القياس، وعلى الرغم من عدم وجود اتفاق كبير على ماهية الأبعاد المسهمة للاتجاهات، إلا أن هناك عددًا من الأبعاد التى اتضحت أهميتها أشرنا إليها فيما سبق، كما أن هناك جهودًا متزايدة تهدف للتوصل إلى اتفاق حول الأبعاد الأساسية للاتجاهات.

(٣) ترجمة هذه الأبعاد النظرية وما تتضمنه من عناصر إلى مواقف يمكن ملاحظتها بحيث يُستند إليها في التوصل إلى أدلة Indicators تتعلق بهذه الأبعاد. وعلى الرغم من اختلاف الإستراتيجيات والأساليب التى تتبع في إجراء ذلك، إلا أن الاستراتيجية أو الأسلوب المستخدم يُعد بمثابة تحديد إجرائي للاتجاه الذى تمثله الملاحظات الإمبيريقية التى تُستمد منها أدلة الاتجاه. وهذا التحديد الإجرائي للاتجاه ربما يتفق أو يختلف إلى حد ما عن التصور المبدئي لطبيعة الاتجاه المراد قياسه، غير أن هذا يعتمد على كفاية الأدلة التى تم اختيارها، فكلما كانت الأدلة تُمثل أبعاد الاتجاه تمثيلا جيدًا يكون الاتفاق بينهما أفضل. ولكن نظرًا لأن هذه الأدلة تكون عادة بمثابة عينات من السلوك الذى ينطوى عليه الاتجاه، فإنه ينبغي توقع بعض الاختلاف بينهما، ويمكن العمل على جعل هذا الاختلاف أقل ما يمكن بمراعاة أن تكون هذه العينات ممثلة للنطاق السلوكي لهذه الأبعاد.



(٤) ضم الدرجات المستمدة من هذه الأدلة فى مؤشرات Indices تُمثل الاتجاه الذى تنطوى عليه. وتُعد هذه المرحلة من المراحل التى تتعلق بدرجة أساسية بمفهوم القياس، أى تعيين أرقام تدل على الملاحظات وفقًا لقواعد محددة. فاختيار أدلة معينة Indicators تُميز الاتجاه المراد قياسه يتطلب تحديد السلوك الملاحظ المتعلق بهذه الأدلة. أى ينبغى تحديد طبيعة الموقف الاختبارى الذى تجمع فيه الملاحظات، وما إذا كان سيقتصر على مواقف اصطناعية يجيب فيها الفرد لفظيًا عن مجموعة من الأسئلة المتعلقة بالاتجاه المعين، أو يجيب تحريريًا عن فقرات استبيان، أو تُجمع الملاحظات فى مواقف حياتية فعلية بعد تحديد جوانب الأدلة التى ينبغى ملاحظتها فى ضوء طبيعة الاتجاه وأبعاده. ويتطلب أيضًا تحويل الملاحظات إلى مقادير كمية وضم هذه المقادير معًا لتصبح درجات على ميزان، واقتراح عمليات تنظيم وتصنيف لهذه الملاحظات بحيث يُعبر عنها النظام التصنيفى المقترح، وربما تُستبعد بعض الملاحظات التى لا تطابق هذا النظام. لذلك يعد النظام التصنيفى مكملاً للتصور المبدئى لطبيعة الاتجاه الذى تحدَّد فى المرحلة الأولى. وتحقيق ذلك فى قياس الاتجاهات يعتمد على نماذج فى بناء الموازين Scaling Models تتطلب تحقق فروض معينة فى الملاحظات أو البيانات لكى نستطيع التوصل إلى خصائص قياسية يفضَّل توافرها. ولكن هذه الخصائص ربما لا تناظر التصور المبدئى لطبيعة الاتجاه المراد قياسه. ويرى كومبس Coombs أنه ربما يصعب تحقق فروض هذه النماذج فى البيانات المتعلقة بالاتجاهات وغيرها من الجوانب الوجدانية.

لذلك يقترح التحرر من فروض هذه النماذج إذا لم تحتمل البيانات هذه الفروض، وفى هذه الحالة ينبغى أن نسمح لخصائص البيانات أن تبرز أثناء عملية التحليل. والحقيقة أنه لا يوجد تعارض بين التوجهين حيث إنهما يختلفان فقط فى القدر من التباين الذى يمكن إعزاؤه إلى الأخطاء العشوائية، والأخطاء التى يمكن احتواؤها فى إطار النموذج المستخدم. لذلك تنقسم هذه النماذج إلى نماذج صارمة فيما تتطلبه من تحقيق فروض معينة فى البيانات، ونماذج لا تتطلب إلا قدرًا ضئيلاً من الفروض. وبالطبع تؤدى النماذج الثانية إلى مؤشرات للاتجاهات أقل دقة من الناحية السيكومترية، ولكن تفيد من جميع البيانات أو الملاحظات التى تم جمعها.

وعمومًا، فإن استخدام أى من هذه النماذج يعتمد بدرجة أساسية على طبيعة البيانات ومستوى قياسها، أى ما إذا كان مستوى القياس اسميًا أو رتبيًا أو فتريًا كما أوضحنا فى الفصل الأول. فالأعداد فى المستوى الاسمى تُصنف البيانات أو الملاحظات المتعلقة بالاتجاه المطلوب إلى أقسام منفصلة، ولذلك فإن هذه الأعداد لا تكون مرتبة وبالتالى يصعب قياس الاتجاهات بمثل هذه الأعداد، أما فى المستوى

الرتبي، فإن الفرد الذي يحصل درجة ولتكن ٤ يُفترض أن اتجاهه أكثر إيجابية من الفرد الذي حصل على الدرجة ٣ مثلاً. فالأعداد في هذا المستوى تكون مرتبة، ولكن المسافات فيما بينهما غير متساوية، فالفرق بين الدرجة ٤ والدرجة ٣ لا تساوي في إيجابية الاتجاه الفرق بين الدرجتين ٦ ، ٥ مثلاً وذلك لعدم توافر صفر مطلق. أي أن الميزان الرتبي يتعلق فقط بترتيب الأفراد وفقًا لدرجة إيجابية الاتجاه لديهم، ولكن الفروق بين الرتب المختلفة لا تكون متساوية، بينما يحدث ذلك في المستوى الفترى حيث تكون الفترات التي تفصل بين الأعداد المتتالية متساوية. فالفرق بين الدرجتين ٤ ، ٣ في إيجابية الاتجاه تساوي الفرق بين الدرجتين ٦ ، ٥ في الإيجابية، ولكن لا نستطيع اعتبار الدرجة ٦ ضعف الدرجة ٣ في الإيجابية نظرًا لعدم وجود صفر مطلق أيضًا، حيث إن هذا يقتصر على المستوى النسبي. غير أننا ذكرنا فيما سبق أن الاتجاهات تتراوح في شدتها بين الإيجابية المرتفعة والسلبية المرتفعة. وهذا يعني أن هناك نقطة محايدة تناظر الصفر على متصل الاتجاه، يتغير عندها الاتجاه من الإيجابية إلى السلبية أو العكس، وتحديد موقع هذه النقطة يعد أمرًا ضروريًا في قياس الاتجاهات. إذ ربما نود معرفة ما إذا كان تعديل اتجاه الفرد من موقع معتدل إلى موقع إيجابي أو سلبي متطرف أيسر من تعديل اتجاهه من موقع يدل على اتجاه إيجابي منخفض إلى حد ما إلى موقع يدل على اتجاه سلبي منخفض، وهل الموقع المحايد يعني التطابق مع الجماعة. والحقيقة أن نقطة الصفر على متصل الاتجاه تعد من مشكلات قياس الاتجاهات. ويرى بعض علماء النفس أن الاتجاهات تكون عادة إما موجبة أو سالبة، فإذا لم تكن كذلك، أي إذا كانت محايدة أو تقع عند نقطة الصفر، لا نستطيع أن نطلق عليها مصطلح الاتجاهات Attitudes.

كذلك فإن تساوي وحدات مُتصل الاتجاهات المُتمثل في المستوى الفتري يُعد أمرًا مطلوبًا، وذلك لكي نتمكن من إيجاد متوسط درجات فرد في موازين مختلفة، ومقارنة التغيرات التي تحدث في اتجاهات الأفراد على هذا المتصل، وكذلك مقارنة درجات أحد الأفراد بدرجات فرد آخر في ميزان معين. غير أن الموازين التي تُستخدم في قياس الاتجاهات، والتي سنوضحها بعد قليل ربما لا تحقق ذلك، وإن كانت وحدات بعض هذه الموازين تفي بذلك إلى حد ما.

وكثير من الموازين التي تقيس الاتجاهات تكون أحادية البُعد Unidimensional، أي أنها تقيس متغيرًا واحدًا أو اتجاهًا واحدًا حيث يمكن تمثيل درجة الفرد بنقطة على متصل الاتجاه. فإذا حصل فردان أو أكثر على درجة متساوية في الاتجاه الذي يقيسه الميزان Attitude Scale ، فإن درجة كل منهم يكون لها نفس الموقع على متصل

الاتجاه، أى يكون لهم نفس الاتجاه تقريبًا. ولكن إذا كان الميزان يقيس أكثر من اتجاه واحد، أى متعدد الأبعاد Multidimensional، فإن الدرجة نفسها يمكن أن تُمثل أنماط اتجاهات مختلفة. فمثلاً إذا كان ميزان اتجاه نحو أفراد المهن المختلفة يشتمل على ست مفردات تتعلق بالاتجاه نحو كل مجموعة، فإن هذا الميزان لا يكون أحادى البُعد، أما إذا كانت موازين الاتجاهات نحو المجموعات المختلفة مرتبطة فيما بينها ارتباطًا مرتفعًا، فإنه يمكن اعتبارها تمثل ميزانًا أحادى البُعد.

وتُعد خاصة البُعد الأحادى Unidimensionality من الخصائص المهمة التى ينبغى توافرها فى الموازين التى تقيس الاتجاهات نظرًا لسهولة تفسير نتائجها، ويعتمد ذلك اعتمادًا كبيرًا على الأسلوب الذى يُستخدم فى بناء الميزان. فإذا كان لدينا أكثر من ميزان واحد يشتمل كل منها على عينة من الفقرات المختلفة المستمدة من نطاق سلوكى واحد للاتجاه المراد قياسه، فإنه ينبغى أن تكون هذه الموازين مرتبطة ارتباطًا مرتفعًا، بل وربما نود أيضًا أن يكون توزيع درجات كل منها يتخذ نفس الشكل العام، وكذلك نفس نمط الارتباطات الداخلية بين فقرات كل منها، ونفس مقدار الثبات.

تصنيف مقاييس الاتجاهات :

تهدف جميع مقاييس الاتجاهات للتوصل إلى استدلالات تتعلق بالاتجاهات استنادًا إلى أدلة يمكن ملاحظتها. وعلى الرغم من تعدد طرق تصنيف هذه المقاييس، إلا أنه تجمعها عناصر مشتركة، إذ يمكن التمييز بين ثلاثة أنواع من النطاقات السلوكية للاتجاه Attitude Universes التى يُطلب فيها من الفرد أن يستجيب استجابة لفظية، وهذه الاستجابة تعبِّر عادة عن رأى الفرد أو تصف مشاعره. ونطاقات هذه الاستجابات يمكن تسميتها الاتجاهات اللفظية المستدعاة Elicited Verbal Attitudes. ولكن تعبير الفرد عن آرائه فى بعض الأحيان فى محادثته المعتادة مع أصدقائه يُمثل الاتجاهات اللفظية التلقائية Spontaneous Verbal Attitude، أما القسم الثالث من النطاقات السلوكية فيتعلق بالسلوك اللفظى أو غير اللفظى أى الفعل المُوجه نحو موضوع الاتجاه، ويمكن تسميتها الاتجاهات المُتمثلة فى الفعل Action Attitudes. وتتمثل هذه الأقسام الثلاثة فى أن يُطلب من الفرد إبداء رأيه فى قضية معينة، أو أن يوضح تلقائيًا لأصدقائه رأيه فى القضية، أو أن يسلك بطريقة معينة تجاهها.

ويمكن تعيين أرقام تناظر الاستجابات فى كل من هذه الأقسام وفقًا لقواعد معينة. وعلى الرغم من أن معظم الموازين المستخدمة تقيس الاتجاهات اللفظية المستدعاة، أى القسم الأول، إلا أنه يمكن بناء أدوات لقياس الاتجاهات اللفظية فى

القسمين الآخرين. إذا يمكن عن طريق أساليب الملاحظة والمقابلة الشخصية التى يعتنى بتصميمها إعداد موازين لقياس الاتجاهات اللفظية التلقائية، والاتجاهات المتمثلة فى الفعل. وإلى جانب هذه الأساليب المباشرة في قياس الاتجاهات، توجد أيضًا أساليب غير مباشرة أو مستنكرة Disguised Techniques، مثل الأساليب الإسقاطية بأنواعها المختلفة، كما توجد أيضًا أساليب فسيولوجية.

لذلك فإن مقاييس الاتجاهات يتم تصنيفها عادة على أساس نوع الأدلة Indicators التى يُستدل منها على الاتجاه المطلوب كالتالى:

القسم الأول: يشتمل على الطرق التى تُمكننا من الاستدلال استنادًا إلى استجابات الفرد وفقًا لمشاعره الحقيقية لمجموعة من الفقرات أو الصفات، وتُعرف هذه الطرق بأساليب بناء الموازين Scaling Techniques، والأدوات التي يتم بناؤها تسمى «موازينScales» . ومن هذه الأساليب الشائعة الاستخدام موازين ثيرستون Thurstone Scales، وموازين ليكرت Likert Scales ، وموازين جتمان Guttman Scales، وبعض الأساليب الخاصة الأخرى مثل أسلوب التمايز اللغوى Semantic Differential Scales الذى اقترحه أوزجود ومعاونوه .Osgood et al، وأسلوب الفَرْد Unfolding Technique الذى اقترحه كومبس Coombs، والأسلوب Q الذى اقترحه ستيفنسون Stephenson، والموازين متعددة الأبعاد Multidmensional وغيرها. وتختلف جميع هذه الموازين عن الاختبارات التى تقيس الجوانب المعرفية فى أن نتائجها لا تُستخدم فى تحديد نجاح الفرد أو إخفاقه، أو قوته أو ضعفه، وإنما تُستخدم فى قياس درجة ما يمتلكه الفرد من الخاصة الوجدانية، أى الاتجاه نحو موضوع معين، مثل اتجاه طلبة المدارس الثانوية نحو التعلم الذاتى، أو اتجاه طلبة الجامعات نحو التعليم المستمر. لذلك لا توجد إجابة صحيحة وإجابة خطأ للفقرات التى يشتمل عليها ميزان الاتجاهات Attitude Scale.

القسم الثانى: يشتمل على الطرق التى تُمكننا من الاستدلال استنادًا إلى سلوك ظاهر Overt. فملاحظة السلوك في مواقف حياتية فعلية يزيد من صدق هذا الاستدلال. وهذه المواقف لا تتطلب تعاون المستجيب، كما لا يجعله يشعر بأنه فى موقف اختبارى أو أنه يُعامل بطريقة خاصة، وكذلك لا تؤدى إلى تغير فى الظاهرة التى يجرى قياسها. ويرى سشرست (Sechrest, 1969) أنه لا ينبغى الاستناد إلى موازين الاتجاهات استنادًا تامًّا، وأوصى باستخدام أساليب ملاحظة السلوك إلى جانب هذه الموازين، فقد بيَّنت كثير من الدراسات التى قام ويكر (Wiker , 1969) بمراجعتها أن

هناك احتمالاً كبيراً أن ترتبط الاتجاهات التى تُقاس بهذه الموازين بالسلوك الظاهر ارتباطًا منخفضًا. واستخلص من هذه الدراسات أن ١٠٪ على الأكثر من تباين درجات مقاييس السلوك الظاهر يمكن تفسيرها باستخدام البيانات المتعلقة بالاتجاهات. ومن أمثلة أساليب ملاحظة السلوك: لعب الأدوار Role Playing ، والاختيارات السوسيومترية Sociometric Choices وغيرهما.

القسم الثالث: يشتمل على الطرق التى تُمكننا من الاستدلال استنادًا إلى مستويات استجابة أكثر عمقًا من الاستجابة لفقرات موازين الاتجاهات. وتُقدَّم المثيرات بحيث يفهمها الفرد المستجيب بطريقة تختلف عما يقصده الفاحص، أى أن هذه الطرق غير المباشرة لقياس الاتجاهات تحمل معنى معينًا بالنسبة للمستجيب، بينما يستند تفسير الاستجابات إلى نظام مختلف فى المعنى، ومن أمثلتها الطرق الإسقاطية كما سبق أن أشرنا.

القسم الرابع: يشتمل على الطرق التى تُمكننا من الاستدلال استناداً إلى مقاييس فسيولوجية للاتجاهات. وقد اقتُرح عدد من هذه المقاييس، مثل الاستجابة الجلفانية للجلد، واتساع أو ضيق حدقة العين، ومعدل نبض القلب عند عرض تمثيل مصور لمواقف تتعلق بموضوع الاتجاه. وعلى الرغم من أن هذه المقاييس تكشف عن الاستثارة أو الشدة، إلا أن المعلومات التى نستمدها منها لا تتعلق تعلقًا تامًا بقياس الاتجاهات، حيث إن تَوجُّه الاستثارة وهدفها يظل غير معلوم. وعلى الرغم من أن فيشباين ، وأجزن (Fishbein & Ajzen , 1965) اقترحا أهمية الحصول على قياس غير لفظى للاتجاهات يكون خارجًا عن تحكم المستجيب، إلا أنهما علّقا على ذلك بأنه ليس من المحتمل أن أيًا من مقاييس رد الفعل الفسيولوجى يمكن أن تحقق هذا الاقتراح.

ونظرًا لأن الطرق التى تنتمى إلى القسم الأول المتعلقة بأساليب بناء موازين الاتجاهات تُعد أكثر استخدامًا من الطرق التى تنتمى إلى القسمين الآخرين، كما يسهل تفسير نتائجها، فسوف نناقشها بالتفصيل فيما يلى، ونوضح بعض الطرق الأخرى بإيجاز.

أساليب بناء موازين الاتجاهات :

أشرنا فيما سبق إلى أنه توجد أساليب متعددة لبناء موازين الاتجاهات Attitude Scales ذكرنا بعضًا منها. وسوف نوضح فيما يلي كيفية استخدام كل من هذه الأساليب، وميزات وعيوب كل منها، ونقارن بينها من حيث الأسس المنهجية التي استندت إليها، ويوضح الشكل التخطيطي التالي (١٢ـ٢) أهم هذه الأساليب:

الأساليب الأساسية في بناء موازين الاتجاهات

أسلوب الفترات المتساوية ظاهريًا	أسلوب التقدير الجمعي	الأسلوب التراكمي	أسلوب تمايز معاني المفاهيم	أسلوب التصنيف الترتيبي (Q)	أسلوب البَسْط
(ثيرستون)	(ليكرت)	(جتمان)	(أورجود)	(ستيفنسون)	(كومبس)
فترى اعتبارى	رتبى	رتبى		تصنيفى	رتبى

شكل (١٢ـ٢) يوضح أساليب بناء موازين الاتجاهات الشائعة الاستخدام

أسلوب الفترات المتساوية ظاهريا (ثيرستون):

يعد هذا الأسلوب من المحاولات المبكرة التى قام بها ثيرستون (Therstone, 1929, 1931) لبناء موازين الاتجاهات، والحقيقة أنه قدم أساليب مختلفة استنادًا إلى تجارب علم النفس الفيزيائي Phsychophysics تُفيد فى بناء موازين لأحكام الأفراد حول خصائص فيزيائية مختلفة مثل الوزن، فكلما كان الفرق الحقيقي بين وزن جسمين ضئيلا قل عدد الأفراد الذين يستطيعون التمييز الصحيح بين وزنيهما، وحاول ثيرستون بطريقة مماثلة بناء موازين اتجاهات بأن يطلب من الأفراد المقارنة بين فقرات أو عبارات تتعلق باتجاه معين مثنى مثنى، والحكم على كل من الأزواج من حيث شدة الاتجاه، أى أيهما أكثر إيجابية وأيهما أكثر سلبية، وأطلق على ذلك أسلوب المقارنات الثنائية Paired-Comparisons. غير أن هذا الأسلوب يصعب إجراؤه إذا كان عدد الفقرات أو

العبارات كبيراً كما هو الحال في معظم موازين الاتجاهات، حيث إن ١٠ فقرات تتطلب إجراء $\frac{١ \times ٩}{٢}$ = ٤٥ مقارنة ثنائية؛ لذلك اقترح ثيرستون أسلوبًا آخر يتطلب إجراؤه جهدًا أقل، وأطلق عليه أسلوب الفترات المتساوية ظاهريًا Equal-Appearing Intervals.

خطوات بناء موازين الاتجاهات باستخدام أسلوب ثيرستون :

يمكن بناء موازين الاتجاهات باستخدام أسلوب ثيرستون باتباع الخطوات التالية:

أولاً: تجميع عدد كبير من الفقرات التي تتعلق بموضوع الاتجاه المراد قياسه، وليكن ١٠٠ فقرة مثلاً تعبِّر عن درجات تفضيل متباينة للموضوع، أي تتراوح بين تفضيل أو موافقة تامة، وعدم تفضيل أو عدم موافقة تامة. ويمكن الاسترشاد في ذلك بمراجعة الأدبيات المتعلقة بموضوع الاتجاه ومن المقابلات الشخصية الاستطلاعية. وينبغي أن تكون صياغة الفقرات واضحة وتعبِّر بدقة عن فكرة واحدة خالية من الغموض، وتسمح بالقبول أو الرفض بدرجات مختلفة فيما يتعلق بموضوع الاتجاه، ولا يجب أن يعبِّر محتواها عن حقائق، وأن تبدو جميع الفقرات مميِّزة، أي تحتمل استجابات مختلفة من الأفراد الذين تختلف شدة اتجاهاتهم نحو الموضوع، وتُكتب كل فقرة على بطاقة صغيرة. فمثلاً إذا كان الاتجاه المراد قياسه نحو التعليم فإن بعض الفقرات ربما تكون كالتالي:

(أ) لا يمكن أن نكون مواطنين صالحين بدون تعليم.

(ب) أذهب إلى المدرسة فقط لأنني مُجبر على ذلك.

(جـ) ينبغي أن يحظى التعليم بمزيد من التمويل.

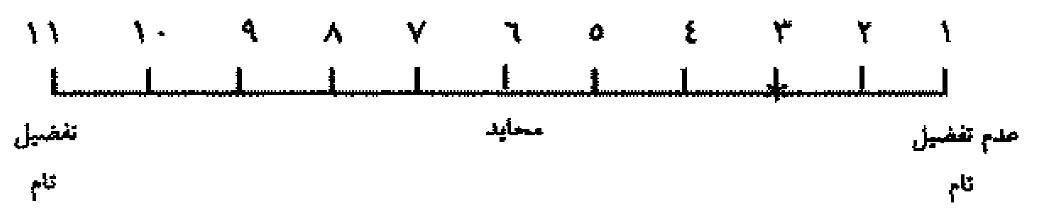
(د) لا يجب أن أبذل جهدًا كبيراً إلا للنجاح في الامتحانات.

(هـ) التعليم مفيد في حياتنا المعاصرة.

(ز) أشك فيما إذا كان التعليم قد أسهم في تطوير الحياة أم لا.

ثانيًا: تُعرض بطاقات الفقرات على عدد من المحكمين الأكفاء، الذين يتراوح عددهم بين ٤٠، ٦٠ محكمًا وربما أكثر، ويتطلب من كل منهم على حدة تصنيفها في مجموعات أو أقسام منفصلة وفقًا لدرجة آرائهم. واقترح ثيرستون أن يكون عددها ٩ أو ١١ مجموعة أو قسمًا، فالمجموعة الأولى يُعيَّن لها الرقم (١) وتدل على درجة مرتفعة من التفضيل، والمجموعة الأخيرة يُعيَّن لها الرقم (١١) مثلا وتدل على درجة منخفضة من التفضيل، بينما الأرقام من (٢) إلى (١٠) تدل على درجات متفاوتة من الأفضلية،

والرقم ٦ يدل على النقطة المحايدة Neutral . وهنا يُفترض أن المسافات بين المجموعات المتتالية تبدو كما لو كانت متساوية من وجهة نظر المحكم. وينبغي مراعاة ألا يقوم المحكمون بتقدير الفقرات من حيث موافقتهم أو عدم موافقتهم على ما تشير إليه كل فقرة، وإنما يضع الفقرة على ميزان أفضلية Favorability Scale من حيث إدراك كل منهم لمدى إيجابية أو سلبية الفقرة فيما تعكسه من الاتجاه الذى يتم تقييمه. فمثلا الفقرة الأولى المبينة أعلاه يفحصها المحكم ويقرر ما إذا كانت تعكس عدم أفضلية وليس عدم أفضلية تامة من حيث الاتجاه نحو التعليم بغض النظر عما إذا كان يوافق أولا يوافق على ما تشير إليه الفقرة، أى أن عليه أن يفكر فقط فى الدرجة التى تعكس بها الفقرة «الأفضلية» لموضوع الاتجاه أى التعليم مستقلا عن اتجاهه نحو الموضوع، ويضع علامة (*) مقابل الرقم المناسب على ميزان الأفضلية كالتالى:



وتكرر العملية لكل فقرة من الفقرات التى تم تجميعها مبدئيًا. وهذا يتطلب أن يكون المحكمون على دراية جيدة بالمفهوم أو الموضوع أو القضية المتعلقة بالاتجاه، وكذلك بردود فعل وعمليات تفكير المجموعة العمرية من الأفراد الذين يُبنى ميزان الاتجاه من أجلهم. وبالطبع لا نتوقّع أن يتفق المحكمون اتفاقًا تامًا على المجموعة أو القسم الذى يرون أن كل بطاقة تنتمى إليه، حيث سنجد تشتتًا أو توزيعًا لأرقام الأقسام المختلفة لكل منها. ويبين هذا التوزيع عدد المحكمين الذين عيّنوا رقمًا معينًا مناظراً لأحد الأقسام ١١، ١٠، ٩ . . . ، ١ لكل فقرة أو بطاقة وليس من الضرورى أن يكون عدد البطاقات متساويًا فى كل قسم.

ثالثًا، نوجد قيمة وسيط Median التوزيع لكل فقرة، وهذه تعد بمثابة قيمة ميزان الفقرة Item Scale Value لأنها تدل على موقع الفقرة على متصل الأفضلية، وبذلك تشير إلى شدة الاتجاه Intensity Scale، وقد سبق أن أوضحنا فى الفصل الثالث كيفية حساب قيمة الوسيط بيانيًا وجبريًا.

رابعًا، نوجد تشتت التوزيع باستخدام مقاييس التشتت التى أوضحناها أيضًا فى الفصل الثالث، ولكن يُفضل استخدام مقياس نصف المدى الربيعى لأنه يتعلق

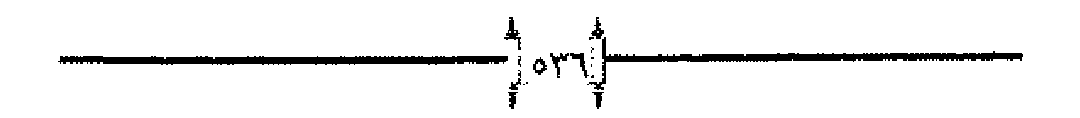
بالوسيط، ويحدد مدى ٥٠٪ الوسطى من الحالات. فإذا كانت قيمة نصف المدى الربيعى لفقرة ما كبيرة، فإن هذا يعنى أن هناك تشتتًا كبيرًا فى الأحكام المتعلقة بفقرة معينة مما يتطلب حذفها نظرا لأنها ربما تبدو غامضة، وبالتالى لا ينبغى تضمينها فى ميزان الاتجاه فى صورته النهائية. وبعد حذف مثل هذه الفقرات تُستبقى الفقرات التى وسيطها أو قيم ميزانها موزعة توزيعًا متعادلاً على المتصل الافتراضى للميزان. وبذلك ربما نحصل على قيم للوسيط مثل ١,٧ ، ١,٨ ، ٢,٢ ، ٢,٢ ، ٢,٩ ، ٣,٠ ، ٣,١ ، ٣,١ ، ٣,٢ ، ٣,٤ ، ٣,٥. فمن المرغوب فيه أن يكون لدينا مجموعة من الفقرات قيم وسيطها تمتد من أحد طرفى المتصل إلى قرب الطرف الآخر بحيث تكون جميع تداريج الاتجاه من الموجب إلى السالب مُمثلة، غير أنه يُحتمل أن نجد عددًا قليلاً من الفقرات متجمعة عند نقط متعددة على طول الميزان، وعندئذ يمكننا اختيار واحدة من الفقرات المتجمعة عند نقطة معينة يكون اتفاق المحكمين عليها متقاربًا. فمثلاً إذا كانت قيمة وسيط كل من فقرتين ٣,١٥ ، ٣,١٩ ووجدنا أن توزيع الأحكام للفقرة الأولى أقل تشتتًا من توزيعها للفقرة الثانية، فإنه ربما نختار الفقرة الأولى ونستبعد الثانية؛ فالفقرتان اللتان وسيط كل منهما ٤ ، ٣ على الترتيب مثلا، يقعان على المتصل ويبتعدان عن بعضهما البعض وحدة واحدة للفرق الذى يمكن ملاحظته Just Noticeable Difference (JND) . لذلك ينبغى اتخاذ قرار فيما يتعلق بمدى المتصل، وعدد الفقرات التى نحتاج إلى اختيارها. وعمومًا فإنه يكفى عادة اختيار ٢٠ أو ٢٥ فقرة، وإن كان من الممكن اختيار عدد أقل. وربما نقرر بناء صيغتين متكافئتين من ميزان الاتجاه إذا كان لدينا عدد كاف من الفقرات يسمح بذلك، ولكن ينبغى أن تُوزع أزواج الفقرات عشوائيًا على كل من الصيغتين بحيث يكون لكل زوج منها نفس قيمة الوسيط ونصف المدى الربيعى، وتُوزع الفقرات عشوائيًّا أيضًا داخل كل صيغة. ويمكن الإفادة من الصيغتين المتكافئتين فى دراسة تغير الاتجاهات.

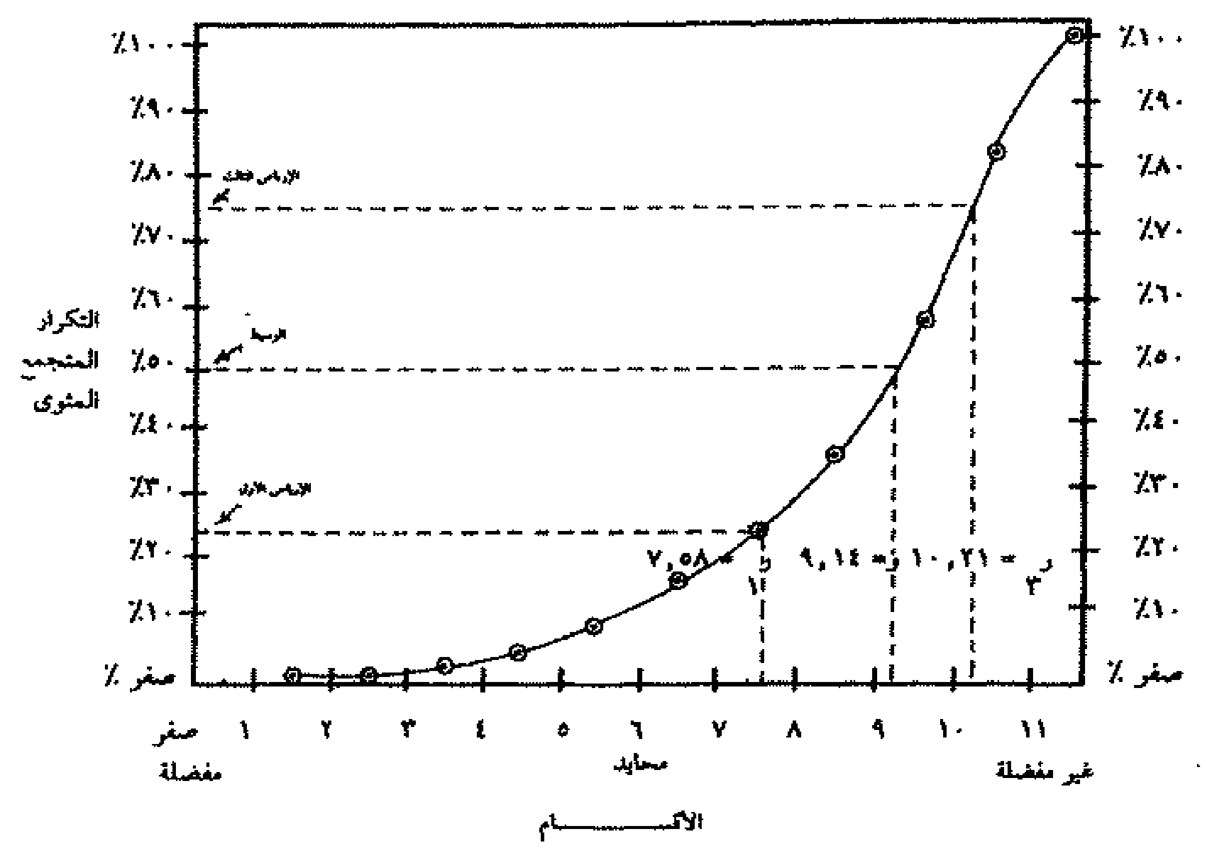
ويمكن توضيح كيفية إيجاد قيم الوسيط ونصف المدى الربيعى لإحدى فقرات ميزان الاتجاه نحو التعليم الذى أشرنا إليه، ولتكن الفقرة الأولى. ونفترض أن توزيع الأحكام كما هو مبين بجدول (١٢-١) التالى:

رقم المجموعة أو القسم	عدد المحكمين	النسب المئوية	النسب المئوية المتجمعة
١	صفر	صفر	صفر
٢	صفر	صفر	صفر
٣	١	٢	٢
٤	١	٢	٤
٥	٢	٤	٨
٦	٤	٨	١٦
٧	٤	٨	٢٤
٨	٦	١٢	٣٦
٩	١١	٢٢	٥٨
١٠	١٢	٢٤	٨٢
١١	٩	١٨	١٠٠
	ن = ٥٠	٪١٠٠	

جدول (١٢ـ١) يوضح توزيع التراضي لبيانات التصنيف لفقرة ميزان الاتجاه نحو التعليم (٣٠٠ محكم)

ويلاحظ من جدول (١٢ـ١) أن المـحكمين الذين بلغ عددهم ٥٠ مـحكمًا، قد صنَّفوا الفقرة في المجموعات أو الأقسام التي عددها ١١ قسمًا، ولم يُصنف أحد من المحكمين الفقرة في القسمين (١) ، (٢) في العمود الأول، أي أن عدد المحكمين (صفر) لكل من هذين القسمين، وحصلنا على النسب المئوية لعدد المحكمين في العمود الثالث بقسمة عدد المحكمين في كل قسم على العدد الكلي للمحكمين وضرب الناتج في ١٠٠. والنسب المئوية المتجمعة في العمود الرابع حصلنا عليها بالجمع المتتالي للنسب المئوية المناظرة لكل قسم في العمود الثالث. ويمكن رسم المنحنى المتجمع الصاعد النسبي بتمثيل الأقسام على المحور الأفقي، والنسب المئوية المتجمعة على المحور الرأسي، ووضع نقط أعلى منتصفات الأقسام تناظر النسب المئوية المتجمعة كما هو موضح بشكل (١٢ـ٢) التالي:





شكل (١٢ـ٢) يوضح المنحنى التكرارى المتجمع الصاعد لفقرة ميزان الاتجاه نحو التعليم

ويتضح من شكل (١٢ـ٢) أن قيمة وسيط هذه الفقرة ٩,١٤ ، ونصف المدى الربيعى = $\frac{ر_٣ - ر_١}{٢} = \frac{١٠,٢١ - ٧,٥٨}{٢} = \frac{٢,٦٣}{٢} = ١,٣٢$

ويمكن بالطبع الحصول على هاتين القيمتين جبريًا (سبق توضيح ذلك فى الفصل الثالث) أى أن قيمة ميزان الفقرة = ٩,١٤ ، والتشتت = ١,٣٢

خامسًا: نختار الفقرات بحيث تكون الفترات التى تفصل بينها متساوية، وقيمة نصف المدى الربيعى لها منخفضة، ويمكن اختيار سعة الفترات وحدة واحدة أو نصف وحدة أو غير ذلك. فإذا قررنا اختيار ٢٠ فقرة مثلا، فإنه ربما نختار الفقرات التى تقترب قيمة وسيط كل منها من ٢,٥ ، ٣ ، ٣,٥ ، ٤ ، ٤,٥ مثلاً ، وهكذا ، وبذلك تكون فترات ميزان الاتجاه متساوية ظاهريًا. وربما نود فى بعض الحالات أن يكون هناك تمييز دقيق فى جزء معين من الميزان عن غيره من الأجزاء، وعندئذ يمكن أن نختار الفقرات التى قيمتها متقاربة فى هذا الجزء.

سادساً: تُرتب الفقرات التى تم اختيارها بطريقة عشوائية، وتُدون فى قائمة دون ذكر قيمة ميزان كل منها Scale Values. وعندما يُستخدم ميزان الاتجاهات يُطلب ممن سيُطبق عليهم إبداء موافقتهم أو عدم موافقتهم فقط على كل فقرة، أو اختيار عدد محدد من الفقرات التى تعبِّر عن موافقتهم بدرجة أفضل. وننظر فقط إلى الفقرات التى أبدى كل منهم موافقته عليها، ونحسب متوسط قيم ميزان هذه الفقرات (الوسيط)، وهذه القيمة المتوسطة تحدد موقع الفرد على متصل الاتجاه (التفضيل - عدم التفضيل). ومن الناحية النظرية ينبغى أن يُبدى الفرد موافقته فقط على عدد قليل من الفقرات المتجاورة فى قيمة ميزان كل منها التى تقترب من موقعه الحقيقى على متصل الاتجاه. فإذا ما أبدى عدد كبير من الأفراد موافقته على فقرات غير متجاورة فى قيمة ميزان كل منها، فإن هذا يمكن أن يكون دليلاً على أن الميزان متعدد الأبعاد، وهذا ربما يتطلب تقسيم الميزان إلى موازين فرعية أحادية البُعد. وعلى الرغم من أن هذا الأسلوب فى بناء موازين الاتجاهات يتعلق بدرجة أساسية بتحديد موقع كل فقرة على متصل الميزان الافتراضى، إلا أنه ينبغى الحيطة عند معالجة قيم الميزان كأعداد فعلية أو كوحدات يمكن جمعها أو تبادلها على المتصل الخطى. فنحن هنا لا نتعامل مع فترات متساوية وإنما مع فترات متساوية ظاهريًا Equal-Appearing Intervals، أى أنها وحدات سيكولوجية وليست وحدات عددية، حيث يمكن أن تكون متساوية أو غير متساوية عدديًا. كما أن تقسيم المتصل إلى (١١) قسم يُعد تقسيمًا اعتباريًا. لذلك لا يجب أن نعتقد أن الدرجة (٨) تساوى ضعف الدرجة (٤) فى قيمتها العددية، كما أننا لا نستطيع المقارنة بين الدرجات من ميزان إلى آخر فى بطارية تشتمل على عدد من موازين الاتجاهات.

بعض ميزات وعيوب أسلوب الفترات المتساوية ظاهريا:

يتميز هذا الأسلوب الذى اقترحه ثيرستون Thurstone بأنه يؤدى إلى موازين للاتجاهات تُمكِّن من التمييز بين عدد كبير من الأفراد فى موقع كل منهم على ميزان فتراته متساوية ظاهريًا كما سبق أن أوضحنا. فعندما نوجد متوسط قيم ميزان الفقرات التى وافق عليها كل فرد نستطيع أن نُحدد مواقع مختلفة لاتجاهاتهم، وبذلك يمكن التمييز الدقيق بينهم فيما يتعلق بالاتجاه نحو موضوع معين. ويرى بعض مؤيدى هذا الأسلوب أن اعتماد قيم ميزان كل فقرة Scale Value على اتفاق مجموعة أكفاء من المحكمين الذين يقومون بفحص كل فقرة واستبعاد الفقرات التى لم تحظ بقدر كبير من الاتفاق يُمكننا من استخدام الميزانِ بمزيد من الثقة فى ثبات درجاته.

غير أنـه قد وُجِّه كثير من النقد لهـذا الأسلوب، وبخاصة فيما يتعلق بالوقت والجهد الذي يتطلبه فـي بناء موازين الاتجاهات من حيث التجميع المبدئي لعدد كبير من الفقرات وتقسيمها بواسطة عدد كبير أيضًا من المحكمين كل على حدة، وحساب قيم ميزان كل فقرة، ومقدار تشتت هذه القيم استنادًا إلى هذا التقييم، واختيار الفقرات التي سوف يشتمل عليها الميزان في صيغته النهائية، وفقًا لمحكات معينة ، وإيجاد متوسط درجات كل فرد يطبق عليه الميزان لإيجاد موقعه النسبي على متصل الاتجاه موضع القياس. كذلك من المحتمل أن يتحدد موقع واحد على المتصل لفردين متوسط قيم ميزان الفقرات التي أبدوا موافقتهم عليها متساويًا على الرغم من اختلاف قيم وسيط الفقرات في الحالتين، إذ يمكن أن يوافق أحدهما على فقرات قيمة وسيط كل منها ٢، ٤، ٦ على الترتيب ومتوسطها ٤، ويوافق الآخر على فقرات مختلفة وسيط كل منها ٣، ٤، ٥ ومتوسطها ٤ أيضًا، مما يجعل تفسير درجة اتجاه كل منهما موضع تساؤل، نظرًا لاختلاف نمط اتجاه كل منهما.

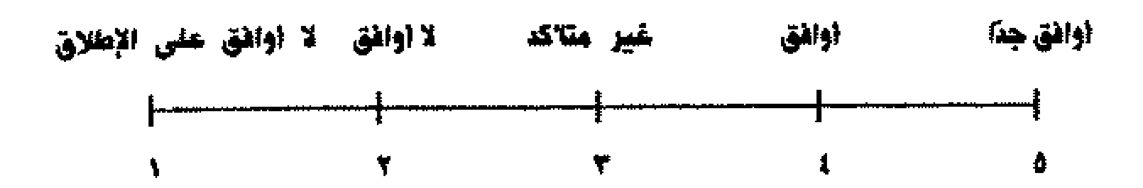
كما أن تصنيف المحكمين للفقرات ربما يتأثر باتجاهاتهم الشخصية وخلفياتهم الثقافية مما يؤدي إلى اختيار فقرات لا تُمثل مواقع ثابتة على متصل الاتجاه. إذ إن أسلوب ثيرستون يستند إلى افتراض أساس يتعلق بعدم اختلاف قيم ميزان الفقرات باختلاف عينات المحكمين. غير أن بعض الدراسات أوضحت أن هذا الافتراض يتحقق إذا لم يكن هناك تحيز واضح من جانب المحكمين فيما يتعلق بمتصل الاتجاه المراد قياسه. أما إذا كان المحكمون متأثرين بشكل ما بموضوع الاتجاه فإن هذا ربما يؤدي إلى تغيُّر في قيم ميزان الفقرات (Ager & Dawes, 1965).

وعمومًا فإن أسلوب الفترات المتساوية ظاهريًا الذي اقترحه ثيرستون لا يتنبأ بالسلوك بدرجة أفضل من غيره من الأساليب الأخرى التي سنوضحها.

أسلوب التقدير الجمعي (ليكرت):

الأسلوب الثاني لبناء موازين الاتجاهات يسمى أسلوب التقدير الجمعي Summated Ratings، أو أسلوب ليكرت Likert Technique. وهذا الأسلوب يتطلب جهدًا ووقتًا أقل مما يتطلبه أسلوب ثيرستون، ويؤدي إلى نتائج مماثلة،حيث إنه لا يعتمد على تقييم المحكمين، لذلك يُعد من الأساليب الشائعة الاستخدام في القياس والبحوث التربوية والنفسية والاجتماعية. ويعتمد هذا الأسلوب على القياس الرتبي للاتجاهات، حيث يقدَّم للفرد قائمة تشتمل على عبارات أو فقرات،ويُطلب منه إبداء

موافقــته أو عدم موافـقته بدرجات مـتفاوتة تعكس شـدة اتجاهه،فالأفـراد الذين تكون موافقتــهم أقل شدة، والأفراد الذين يتباين مـوقعهم على هذا المتــصل يتوقع أن يكون سلوكهم مـختلفًا فى المواقف الاجـتماعية المتـعلقة بموضوع الاتجاه. غـير أن ارتباط السلوك الفـعلى بالاتجاه لا يزال مـثار جدل بين علمـاء النفس، وتتحدد شـدة الاتجاه جزئيًا بإعطـاء أوزان مختلفة لاستجابة الـفرد لكل فقرة. فمثلاً إذا كـانت الفقرة «أحب مهنتى» فإنه يمكن أن يستجيب الفرد على مـيزان رتبى متدرج يشتمل على خمس نقاط كالتالى:



فالاستجابة «أوافق جدًا» يُعيَّن لها الوزن الرقمى (٥)، والاستجابة «لا أوافق على الإطلاق» يُعيَّن لهـا الوزن الرقمى (١) وهكذا. وكلما زاد الوزن الـرقمى دل ذلك على حب الفرد لمهنته والعكس صحيح، فـإذا استجاب ١٠٠ فرد على هذه الفقرة، فسوف نحصل عـلى ١٠٠ استجابة تـتراوح أوزانها الرقمـية بين (١) ، (٥). فجمـيع الأفراد الذين يستجيـبون «أوافق جدًا» يحصلون على الوزن الرقمى (٥)، أما الذين يـستجيبون «لا أوافق على الإطلاق» يحـصلون على الوزن الرقمى (١). ولكن لـيس بالضرورة أن جميع من أبدوا مـوافقتهم التامة على الفقـرة يكون لديهم نفس المشاعر نحـو مهنتهم؛ لأن فقرة واحـدة لا تكفى بالطبع لتقـييم هذه المشاعـر. لذلك يستند أسلوب الـتقدير الجمعى على مجموعة من الفقرات من أجل التـوصل إلى تقييم أكثر دقة، والتمييز بين الأفراد فى شدة اتجاههم نحو موضوع معين عـلى ميزان متدرج مثل المبين أعلاه. فإذا كان المـيزان يشتـمل على خمس نـقاط، وكان لدينـا (١٠) فقرات، فإن أعـلى درجة يحصل عـليها فرد سـوف تكون (٥٠) وأقل درجة (١٠)، أى أن المـدى أصبح (٤١) للفقرات العشر بدلاً من (٥) للفقـرة الواحدة، وكلما زاد عدد الفقرات زاد المدى، أى زاد تباين الدرجات مـما يؤدى إلى زيادة دقة القياس. وبالطبع لا يقـتصر هذا الأسلوب على نمط الاستـجابات الموضح أعلاه، وإنمـا يمكن استخدام أى ميـزان رتبى متدرج آخر، مثل:

(١) موافق، غير متأكد ، غير موافق.

(٢) موافق بشدة، موافق، غير متأكد ولكن موافق إلى حد ما، غير متأكد ولكن غير موافق إلى حد ما، غير موافق، غير موافق على الإطلاق.

(٣) دائمًا ، غالبًا، أحيانًا، نادرًا، إطلاقًا.

ويُفضل ألا يزيد عدد هذه الأقسام عن خمسة لكي يتمكن الفرد من التمييز بينها واختيار درجة موافقته بدقة. كما يلاحظ أن النمطين ١ ، ٢ اشتملا على القسم «غير متأكد» حيث يُفترض أن الفرد الذى يختاره لا يكون له رأى، ومع هذا يستخدم أحيانًا فى مثل هذه الموازين. ولكن النمط الثانى اشتمل على ستة أقسام لكى يحدد الفرد توجهه سواء جهة اليمين إذا اختار «غير متأكد ولكن غير موافق إلى حد ما» أو جهة اليسار إذا اختار «غير متأكد ولكن غير موافق إلى حد ما»، وذلك لأن الاستجابة «غير متأكد» لا تفيد فى تحليل البيانات. وبالطبع فإن استخدام القسم «غير متأكد» فى موازين الاتجاهات يعتمد على خصائص عينة المستجيبين ومستوى تعليمهم ومدى ألفتهم بالمطلوب منهم.

خطوات بناء موازين الاتجاهات باستخدام أسلوب ليكرت :

يمكن بناء موازين الاتجاهات باستخدام أسلوب التقدير الجمعى (أسلوب ليكرت) باتباع الخطوات التالية:

أولا : تجميع عدد كبير من الفقرات التى تتعلق بموضوع الاتجاه المراد قياسه مثل الاتجاه نحو المعلم، أو نحو مجال دراسى معين، أو نحو موضوعات عامة مثل الانفتاح الاقتصادى، أو تقنيات الاتصال وغير ذلك؛ ونظرًا لأن كل فقرة فى الميزان التجمعى تهدف للحصول على تباين متسق فيما يتعلق بالاتجاه، فإن معظم الفقرات ينبغى أن تتضمن مدى متسعًا من حيث شدة التفضيل Favorablness ، مع مراعاة أنه لا توجد فقرات محايدة فعلاً فى مثل هذا الميزان أو فقرات متطرفة . والفقرات المتطرفة فى الشدة فى كلا الاتجاهين تؤدى إلى تباين أقل من الفقرات الأقل تطرفًا، ويكون تمييزها ضعيفًا، ويمكن أن يتراوح عدد الفقرات بين ٤٠ ، ٦٠ فقرة تكون مقسمة تقسيمًا متعادلا بين الفقرات الموجبة والفقرات السالبة.

ثانيًا : تطبيق الفقرات على مجموعة من الأفراد مماثلة للمجموعة المستهدفة أى التى سيُطبق عليها ميزان الاتجاه فى صيغته النهائية فيما بعد، ويمكن أن تشتمل هذه المجموعة على عدد من الأفراد يساوى عشرة أمثال عدد الفقرات (Nunnally, 1978). ويفضل زيادة عدد الأفراد عن ذلك إذا أمكن،ويُراعى جعل شروط أو ظروف التطبيق

مماثلة بقدر الإمكان لظروف التطبيق على المجموعة المستهدفة، ويُطلب من كل فرد أن يُبدى رأيه فيما يتعلق بكل فقرة ويُحدد شدة موافقته، أى ما إذا كان موافقًا جدًا، أو موافقًا، أو غير متأكد، أو غير موافق، أو غير موافق على الإطلاق؛ لذلك ينبغى أن تكون مجموعة الأفراد ممثلة للمدى الكلى للاتجاه المراد قياسه بقدر الإمكان.

ثالثًا: تُعيَّن درجة أو وزن رقمى لاستجابة الفرد على كل فقرة وفقًا للأوزان التى أشرنا إليها، مع مراعاة أن تُعكس هذه الأوزان الرقمية للفقرات السالبة .

فمثلاً: إذا كانت الفقرة كالتالى: «تُعلّمنى المدرسة ما يساعدنى فى الحصول على عمل» فإن: موافق جدًا = ٥ ، وموافق = ٤ ، وغير متأكد = ٣ ، وغير موافق =٢ وغير موافق على الإطلاق = ١ .

أما إذا كانت الفقرة كالتالى : «المدرسة مضيّعة للوقت» فإن : موافق جدًا = ١، وموافق = ٢ ، وغير متأكد = ٣ ، وغير موافق = ٤ ، وغير موافق على الإطلاق = ٥ .

وتجمع درجات الفرد فى جميع الفقرات التى استجاب لها، فمثلاً إذا استجاب «موافق جدًا» للفقرة الأولى (الدرجة = ٥) ، «وغير متأكد» للفقرة الثانية (الدرجة =٣)، «وموافق» للفقرة الثالثة (الدرجة = ٤) ، «وغير موافق» على الفقرة الرابعة (الدرجة=٢)، فإن درجة الفرد = ٥ + ٣ + ٤ + ٢ = ١٤، وبالمثل فى الفقرة الثانية مع مراعاة الأوزان الموضحة.

رابعًا: إجراء تحليل فقرات Item Analysis لاختيار الفقرات التى تميز بين الأفراد بدرجة أفضل على متصل الاتجاه، وذلك بإيجاد الارتباط بين الدرجات الكلية التى يحصل عليها الأفراد واستجاباتهم لكل فقرة، والاستبقاء على الفقرات التى يكون ارتباطها مرتفعًا، وبذلك نحصل على ميزان متسق داخليًا. أو يمكن إيجاد معامل تمييز كل فقرة كما سبق أن أوضحنا فى الفصل السابع. ويفضل ترتيب قيم معامل الارتباط ترتيبًا تنازليًا واختيار عدد متساو تقريبًا من الفقرات الموجبة والفقرات السالبة وضمهما معًا لتكوّن ميزان الاتجاه فى صيغته الأولية.

خامسًا: إيجاد معامل ألفا لكرونباك (α) Cronbach للفقرات التى تم اختيارها، وذلك بإيجاد تباين درجات كل منها وتباين الدرجات الكلية وتطبيق الصيغة (٤-١٢) التى سبق أن أوضحناها فى الفصل الرابع،أو تكوين مصفوفة الارتباطات بين درجات الفقرات والدرجات الكلية وإجراء تحليل المفردات باستخدام الحاسوب.

وسوف تفيد هذه المصفوفة أيضًا في الخطوة التالية. فإذا كانت قيمة هذا المعامل مرتفعة بدرجة كافية (٠,٨٠ أو أكبر) فإنه يمكن أن يشتمل ميزان الاتجاه على هذه الفقرات في صيغته النهائية (حوالي ٢٠ فقرة). وينبغي ملاحظة ما ذكرناه سابقًا من أن قيمة معامل الثبات تعتمد على عدد أقسام ميزان التقدير المقابل لكل فقرة. وهذه الخطوة تعد بمثابة مؤشر لخاصة أحادية البُعد Unidimensionality، حيث نتحقق بها من أن جميع الفقرات تقيس شيئًا واحدًا، وهو ما اهتم به ليكرت اهتمامًا أساسيًا.

سادسًا : ترتيب الفقرات المُختارة بطريقة عشوائية في ميزان الاتجاه، أي الاستبيان، وبذلك يمكن استخدامه في قياس الاتجاه المطلوب حيث يُطبق على المجموعة المستهدفة، وإيجاد الدرجة الكلية لكل فرد بجمع الدرجات المناظرة لاستجاباته على الفقرات التي يشتمل عليها الاستبيان.

والدرجة الخام الكلية التي يحصل عليها الفرد يكون معناها محدودًا في ذاتها، ويتضح معناها بدرجة أفضل إذا قورنت بدرجات الأفراد الآخرين الذين طبق عليهم ميزان الاتجاه، حيث يمكن أن يتضح موقعه على متصل الاتجاه بالنسبة لموقع هؤلاء الأفراد.

مزايا وعيوب أسلوب التقدير الجمعي :

يتميز هذا الأسلوب بأنه يجمع في بناء موازين الاتجاهات بين خبرة من يقوم بإعدادها والتفكير المنطقي في اشتقاق الفقرات من نطاق نظري شامل للاتجاه المراد قياسه. ويمكنه صياغة الفقرات بما يراه مناسبًا مادام ملتزمًا بالربط المنطقي بين محتوى الفقرات وموضوع الاتجاه. ويستند هذا الأسلوب إلى عدد من الفروض أقل مما تستند إليه الأساليب الأخرى، ويتميز بسهولة حساب درجات الفقرات والدرجات الكلية ومقارنتها بغيرها من الدرجات على متصل الاتجاه. ويعتمد تفسير الدرجات على منطق بسيط مؤداه أنه كلما زادت الدرجة الكلية التي يحصل عليها الفرد دل ذلك على زيادة شدة اتجاهه.

لذلك يُستخدم هذا الأسلوب بكثرة في تصميمات البحوث التي تستكشف بعض الظواهر النفسية الاجتماعية، ويتميز كذلك بالمرونة حيث يستطيع الباحث أن يجعل ميزان الاتجاه مشتملاً على عدد من الفقرات يراه مناسبًا، وزيادة عدد الفقرات يمكنه من الكشف عن الفروق الفردية في الاتجاه المراد قياسه، وبخاصة إذا كان عدد أفراد

المجموعة كبيرًا. غير أن بعض الباحثين يطلقون على أى ميزان متدرج يشتمل على عدد من الأقسام «ميزان ليكرت» على الرغم من أنهم لا يستخدمون فى بنائه أسلوب تحليل الفقرات أو التحقق من أن الفقرات تقيس بُعدًا أحاديًا، وبذلك يُساء استخدام هذا الأسلوب، ولا يُفاد من نتائجه الفائدة المرجوة.

ولكن يُعاب على أسلوب التقدير الجمعى أنه يقتصر على مستوى القياس الرتبى، أى يسمح بترتيب الأفراد على متصل الاتجاه من حيث درجة موافقة كل منهم دون معرفة مقدار زيادة درجة موافقة الفرد عن غيره من الأفراد، أو تحديد مقدار التغير الذى حدث لديه فى الاتجاه بعد مروره بخبرات معينة، حيث إن ذلك يتطلب مستوى قياس فترى. كما أن الدرجة الكلية لفرد ما لا يكون لها معنى واضح نظرًا لأن أنماطًا متعددة من الاستجابات للفقرات يمكن أن تؤدى إلى الدرجة الكلية نفسها. غير أنه يمكن التغلب على ذلك بالعناية بتحليل الفقرات بحيث نضمن أن يشتمل الميزان فقط على الفقرات التى تُميز جيدًا بين الدرجات الكلية المرتفعة والدرجات الكلية المنخفضة، وقد سبق أن رأينا أن هذا ينطبق أيضًا بدرجة كبيرة على أسلوب ثيرستون نظرًا لكثرة أنماط الاستجابات الممكنة فى هذه الحالة. غير أن البعض يرى أن اختلاف أنماط الاستجابات التى تؤدى إلى الدرجة الكلية نفسها يمكن إعزاؤه إلى تباين الخطأ فى الاستجابات وليس إلى تباين حقيقى فى الاتجاه. ونظرًا لتعدد الفقرات التى يشتمل عليها الميزان فإن هذه الاختلافات فى أنماط الاستجابات لكل فقرة التى لا تتعلق بالاتجاه الذى يجرى قياسه يلغى بعضها البعض الآخر، ومع هذا فإنه ينبغى الاهتمام بفحص أنماط الاستجابات لكل فقرة أكثر من الاهتمام بالدرجات الكلية نظرًا لأن معناها يختلف باختلاف هذه الأنماط.

الأسلوب التراكمى لجتمان:

إن موازين الاتجاهات التى تُبنى استنادًا إلى الأسلوبين السابقين تشتمل على فقرات غير متجانسة فيما يتعلق بالأبعاد المختلفة لموضوع الاتجاه المراد قياسه. فمثلاً فى بناء ميزان اتجاه متعلق بالتعليم ربما لا نفصل بين الفقرات أو العبارات التى تتعلق بالممارسات السائدة فى التعليم الإلزامى، والفقرات أو العبارات التى تتعلق بفكرة

التعليم الذاتى غير الرسمى، أو غير ذلك من جوانب أو أبعاد الاتجاه نحو التعليم. فالجمع بين جوانب أو أبعاد مختلفة فى ميزان واحد يؤدى إلى صعوبات فى تفسير الدرجات التى يحصل عليها الأفراد تفسيراً واضحاً، وبخاصة أن أنماط هذه الدرجات تكون متعددة كما سبق أن أوضحنا.

لذلك اقترح جتمان Guttman أسلوبًا للتغلب على هذه المشكلة أطلق عليه الأسلوب التراكمى Cummulative Technique، ويُمكن باستخدام هذا الأسلوب بناء موازين اتجاهات أحادية البُعد Unidimensional.ويُعد ميزان الاتجاه أحادى البُعد فقط إذا أدى إلى ميزان تراكمى، ونعني بذلك أن الفقرات ترتبط فيما بينها بحيث إذا وافق الفرد على الفقرة (٢) فإنه يوافق أيضًا على الفقرة (١) التى تسبقها، ومن يوافق على الفقرة (٣) فإنه يوافق أيضًا على الفقرتين (١)، (٢) اللتين تسبقانها، وهكذا.

فمثلاً إذا طلبنا من الأفراد إبداء موافقتهم أو عدم موافقتهم على العبارات التالية المتعلقة بالاتجاه نحو منظمة الأمم المتحدة:

(أ) منظمة الأمم المتحدة تُنقذ جميع الشعوب.

(ب) منظمة الأمم المتحدة هى أملنا فى التعايش السلمى.

(جـ) منظمة الأمم المتحدة تُعد قوة بنَّاءة فى العالم.

(د) ينبغى استمرار المشاركة فى منظمة الأمم المتحدة.

فإذا كانت هذه الفقرات تمثل ميزانًا تراكميًّا، فإننا نتوقع أن الفرد الذى يوافق على الفقرة (أ) سوف يوافق على الفقرات (ب) ، (جـ) ، (د) ، والفرد الذى لا يوافق على الفقرة (أ)، ولكنه يوافق على الفقرة (ب) نتوقع أن يوافق على الفقرتين (جـ) ، (د) ، وبذلك يمكن ترتيب جميع استجابات الأفراد فى نمط يماثل الموضح بجدول (١٢-٢) التالى، حيث تُقدَّر درجة واحدة لكل استجابة «موافق» على أى من الفقرات الأربع:

الدرجة	الاستجابة «موافق» على الفقرة د	جـ	ب	أ	الاستجابة «غير موافق» على الفقرة د	جـ	ب	أ
٤	×	×	×	×				
٣		×	×	×	×			
٢			×	×	×	×		
١				×	×	×	×	
صفر					×	×	×	×

جدول (١٢ـ٢) يوضح نمط الاستجابات على أربع فقرات فى ميزان تراكمى

خطوات بناء موازين الإتجاهات باستخدام أسلوب جتمان :

عند بناء ميزان اتجاه تراكمى ينبغى كتابة أو تجميع عدد من الفقـرات تتطلب الاستجابة «موافق» أو «غير موافق» وترتيبها بحسب درجة الموافقة، وأحيانًا يمكن ترتيبها عشوائيًا دون محاولة تحديد ما إذا كانت الفقرات الفاصلة بينها متساوية، حيث إن الميزان التراكمى يُعد من المستوى الرتبى. ولكن ينبغى التحقق من أن جميع الفقرات تمثل بُعدًا أحاديًا للاتجاه المراد قياسه، أى تتّخذ نمطًا مثلثيًا مماثلاً لما هو موضح بجدول (١٢ـ٢)-فإذا أمكن التوصل إلى هذا النمط التام يصبح لمجموعة الفقرات التى يتم اختيارها معنى واحد، أى إذا علمنا الدرجة التى يحصل عليها فرد يمكن دون الرجوع إلى استبيان الاتجاه معرفة الفقرات التى أبدى موافقته عليها.

ولكن يصعب الحصول فى الواقع الفعلى على مثل هذا النمط المثلثى التام الذى يعكس بُعدًا أحاديًا لميزان الاتجاه، وذلك لأنه يوجد عادة عدم اتساق فى الاستجابات، أى توجد أخطاء . ويمكن توضيح ذلك بنمط الاستجابات الذى يوضحه جدول (١٢ـ٣) التالى:

الفقرات						
		أكثر شدة			أقل شدة	
الفرد	أ	ب	جـ	د	هـ	الدرجة
١	+	+	(−)	+	+	٤
٢	+	(−)	+	+	+	٤
الخطأ	صفر	١	١	صفر	صفر	

جدول (١٢ـ٣) يوضح الخطأ في بعض أنماط الاستجابات

ويتضح من جدول (١٢ـ٣) أن الفرد الأول استجاب بالموافقة على الفقرتين (أ)، (ب) ، ولم يوافق على الفقرة (جـ) ، ووافق على الفقرتين (د) ، (هـ) وحصل على الدرجة (٤).

فعدم موافقته على الفقرة (جـ) يُعد في هذه الحالة من الخطأ Error ، وكذلك لم يوافق الفرد الثاني على الفقرة (ب) حيث يُعد ذلك أيضًا من الخطأ، لذلك وضعنا دائرة على الفقرتين (جـ) ، (ب) بالنسبة لكل من الفردين على الترتيب، وربما يدل ذلك على أن ميزان الاتجاه ليس أحادى البُعد. أما نمط الاستجابات الموضح في جدول (١٢ـ٢) فيدل على ميزان اتجاه تراكمى أحادى البُعد. ويمكن التنبؤ بنمط الدرجات من معرفتنا للدرجة الكلية للفرد حيث إن نمط الاستجابات ليس به أخطاء (علامات سالبة)، أي يمكن إعادة تكوين هذا النمط Reproducable . وإذا كان عدد الأخطاء في نمط استجابات مجموعة من الأفراد كبيرًا، فإن هذا يعنى عدم إمكانية إعادة تكوين هذا النمط Nonreproducable.

وقد اقترح جتمان Guttman صيغة يمكن باستخدامها إيجاد معامل إعادة تكوين نمط الاستجابات Coefficient of Reproducibility كالتالي:

$$\text{معامل إعادة تكوين نمط الاستجابات} = ١ - \frac{\text{عدد الأخطاء}}{\text{عدد الاستجابات}} \quad \ldots\ldots (١٢\text{ـ}١)$$

والمقصود بالأخطاء كما أشرنا الابتعاد عن النمط المثلثي المثالي.

وعدد الاستجابات = عدد الأفراد × عدد الفقرات.

فإذا نظرنا إلى جدول (١٢-٣) نلاحظ أن عدد الأخطاء = ٢، وعدد الأفراد = ٢، وعدد الفقرات = ٢ ، وبتطبيق الصيغة (١٢-١) نجد أن:

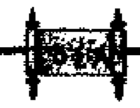
معامل إعادة تكوين نمط الاستجابات = ١ - $\frac{٢}{٢\times٢}$ = ٠,٥٠

وقد اقترح جتمان أن يكون الحد الأدنى لقيمة هذا المعامل لمجموعة من الفقرات التي يشتمل عليها الميزان التراكمي للاتجاه ٠,٩٠ لكي يُطلق على الميزان أحادي البعد. ونظراً لأننا اعتمدنا في المثال السابق على فقرتين فقط انخفضت قيمة هذا المعامل إلى ٠,٥٠ . لذلك أوصى جتمان اختيار ١٠ أو ١٢ فقرة وتُطبق على عدد لا يقل عن ١١٠ فرد، ويتم اختيار هذه الفقرات من نطاق شامل للفقرات المتعلقة بموضوع الاتجاه.

ميزات وعيوب الأسلوب التراكمي:

يتميز هذا الأسلوب كما اتضح لنا بتركيزه على خاصة البُعد الأحادي للفقرات في موازين الاتجاهات، وبذلك يمكن تحديد عدم اتساق استجابات الأفراد، وربما التعرف على الاستجابات غير الصادقة مما يزيد من الثقة في المعلومات التي يقدمها المستجيبون. ويسهل استخدام الأسلوب التراكمي إذا كان عدد الفقرات قليلاً، أما إذا زاد العدد عن ١٢ فقرة، فإن استخدامه يكون مجهداً، بل غير عملي، كما يتميز بأنه يمكن إعادة تكوين أنماط الاستجابات إذا علمنا الدرجة الكلية التي حصل عليها الفرد في ميزان الاتجاه، وهذا بالطبع لا يتحقق في أسلوب ثيرستون أو ليكرت.

غير أنه ليس من الأساليب الأكثر فاعلية في قياس الاتجاهات نحو موضوعات معقدة أو في التنبؤ بالسلوك المتعلق بها. كما أن خاصة البُعد الأحادي ربما تنطبق على مجموعة معينة من الأفراد دون الأخرى ، أو ربما تنطبق عليها في وقت ما ولا تنطبق في وقت آخر، وبخاصة أن التطورات التي حدثت في أساليب بناء موازين الاتجاهات تؤكد أهمية بناء موازين متعددة الأبعاد حيث أوضحت البحوث السيكولوجية أن الأفراد لا يستجيبون عادة استجابات أحادية البُعد، مما يجعل استخدام أسلوب جتمان غير واقعي. كما أن هذا الأسلوب يُعد من الأساليب التي تنتمي إلى النماذج الحتمية في حين أن التوجهات المعاصرة اهتمت ببناء نماذج احتمالية يُستند إليها في قياس السمات الإنسانية بما في ذلك الاتجاهات. ولعل النظرية المعاصرة في القياس التربوي والنفسي التي سنوضحها في الفصل الرابع عشر تُعد أحد هذه التوجهات.



بعض الأساليب الخاصة:

تُعد أساليب بناء الموازين التى ناقشناها فيما سبق أكثر الأساليب استخدامًا فى قياس الاتجاهات، غير أن هناك أساليب خاصة أخرى أقل استخدامًا ولكنها تفيد أيضًا فى قياس كثير من الجوانب الوجدانية للأهداف التعليمية وكذلك فى قياس الاتجاهات.

وسوف نوضح بعض هذه الأساليب فيما يلى:

أسلوب تمايز معانى المفاهيم Semantic Differential Technique :

قام أورجود Osgood بجامعة إيلينوى الأمريكية عام ١٩٥٢ بنشر نتائج دراساته المتعلقة بمكونات معانى المفاهيم، حيث جمع عددًا كبيرًا من الصفات ووضعها فى زواج من الصفات المتضادة ثنائية القطب Bipolar، وطلب من الأفراد تقدير سلسلة من المفاهيم أو الحكم عليها من حيث ما تعنيه لكل منهم باستخدام هذه الأزواج من الصفات. وأجرى تحليلاً عامليًا للبيانات حيث استخلص ثلاثة عوامل أو مكونات أساسية تنطوى عليها هذه التقديرات هى: التقييم Evaluation، والقوة Potency، والنشاط Activity. وفيما يلى بعض أمثلة لأزواج الصفات المتضادة المتعلقة بكل من هذه العوامل:

عامل التقييم: جيد ـ ردىء ، عادل ـ غير عادل ، أمين ـ غير أمين
ناجح ـ غير ناجح ، موجب ـ سالب ، سار ـ غير سار

عامل القوة: قوى ـ ضعيف ، ثقيل ـ خفيف ، كبير ـ صغير
متشدد ـ متساهل ، سميك ـ رفيع ، مذكر ـ مؤنث

عامل النشاط: سريع ـ بطئ ، نشط ـ خامل ، ساخن ـ بارد
مثير ـ هادئ ، متوتر ـ مسترخى ، مندفع ـ مترو

وقد نشر أورجود ومعاونوه (Osgood et al., 1957) نتائج هذه البحوث فى كتاب «قياس المعنى The Measurement of Meaning» وأشاروا إلى كثير من أزواج الصفات المتعلقة بكل من هذه العوامل أو المكونات الأساسية. واستند أورجود فى ذلك إلى افتراض أن خصائص الأفكار والأشياء فى اللغة المكتوبة والمتخاطب بها يتم التواصل بها بمدرجة أساسية بواسطة الصفات. واقترح أسلوبًا عامًا يمكن باستخدامه الحصول على تقديرات للمفاهيم فى سلسلة من موازين للصفات المتضادة ثنائية القطب أطلق عليه «أسلوب تمايز معانى المفاهيمThe Semantic Differential Technique».
و يُعد أسلوبًا لقياس معنى شىء ما بالنسبة للفرد، ويُستخدم بعامة للإشارة إلى أى

مجموعة من موازين التقدير Rating Scales يشتمل كل منها على صفة ثنائية القطب. لذلك فإن هذا الأسلـوب لا يُعد أداة قياس معيـنة أو اختبار، وإنما أسلـوب قياس عام يتميز بـالمرونة فى قياس الاتجاهات وغـيرها من الجوانب الوجدانيـة. ولتوضيح ذلك نفتـرض أن المثير أو الـمفهوم المـراد تقييم اتجـاه الفرد أو مشـاعره نحوه هو مـفهوم «التعلم»، فيمكن أن تكون موازين التقدير التى تشتمل على صفات متضادة ثنائية القطب كما هو موضح بشكل (١٢ـ٣) التالى:

التعلم

سريع	___ : ___ : ___ : ___ : ___ : ___ : ___	بطئ
جيد	___ : ___ : ___ : ___ : ___ : ___ : ___	ردئ
غير مفيد	___ : ___ : ___ : ___ : ___ : ___ : ___	مفيد
سار	___ : ___ : ___ : ___ : ___ : ___ : ___	غير سار
خامل	___ : ___ : ___ : ___ : ___ : ___ : ___	نشط
قوى	___ : ___ : ___ : ___ : ___ : ___ : ___	ضعيف

شكل (١٢ـ٣) يوضح صفحة تمايز معانى المفاهيم

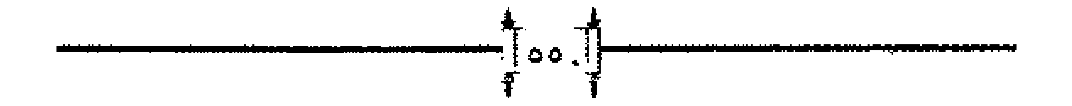
ويُطلب من كل فرد تقديـر اتجاهه نحو المفهوم المدَّون أعلى الصـفحة الخاصة بموازين ذات سبع نقاط يشتمل كل منها على صـفة ثنائية القطب كما هو موضح بشكل (١٢ـ٣)، وأن يضع عـلامة فى المـوضع الذى يعبِّر عن تـوجُّه وشدة مـشاعره نـحو المفـهوم. ويمكن اسـتخدام أى عـدد من المفـاهيم أو المثـيرات مثل: الأشـخاص، والأشياء، والمفاهيم المجردة، والممـارسات، والمؤسـسات وغير ذلك. وبذلك يمكن تقييم اتجاه الأفراد نحو عدد كبير من المفاهيم أو العبارات فى وقت قصير بدلاً من بناء ميزان اتجاه لكل مفهوم على حدة كما فى الأساليب التى سبق أن ناقشناها.

وقد أوصت كـثير من الدراسـات بالاهتمـام بعامل أو مكونـة «التقيـيم» فى بناء الموازين باستخدام أسلوب تمايز معانى المفاهيم.

خطوات أسلوب تمايز معانى المفاهيم:

يمكن اتباع الخطوات التالية عند بناء موازين تمايز معانى المفاهيم:

(١) تحديد المفهوم أو المفاهيم المراد تقديرها: يعتمد اختيار عدد ونوع المفاهيم على الهدف من بناء ميزان الاتجاهات، ويفـضل اختيار مجموعة من المفاهيم



المترابطة التي يجمعها إطار مشترك، فكلما زاد تجانس المجموعة أمكن اختيار صفات ثنائية القطب واضحة المعنى. فمثلاً مجموعة من المفاهيم مثل: التعلم، والمعلم، والمدرسة، والدراسة، والكتب المدرسية ـ تُعد مجموعة متجانسة نسبيًا، ويمكن للمعلم تقييم مشاعر الطلاب نحو مفاهيم أساسية في وحدة أو مقرر دراسي معين.

(٢) اختيار موازين ثنائية القطب مناسبة : يعتمد اختيار مجموعة من الموازين على مدى مواءمتها وتمثيلها للمفاهيم المحددة، وينبغي عدم اختيار صفات طبيعية أو فيزيائية للأشياء، فمثلا لا يجوز إذا كان المطلوب تقييم مفهوم «صخرة» مثلا أن نختار صفة ثنائية القطب مثل (خشن ـ أملس). ويجب أن تكون الموازين ممثلة لمدى متسع من المشاعر التي يحتمل أن تقترن بالمفاهيم المحددة، وربما يتطلب ذلك تجريب هذه الموازين ميدانيًا وإجراء تحليل عاملي للبيانات، غير أن ذلك ربما لا يكون مناسبًا في كثير من الحالات. ويمكن الاسترشاد في اختيار الصفات بالقوائم التي أعدها أوزجود Osgood في كتابه الذي أشرنا إليه، وكذلك اقتراح صفات أخرى مناسبة. وينبغي مراعاة أن يكون مستوى صعوبة الكلمات التي تدل على صفات مناسبًا للمختبرين ومألوفًا لهم.

(٣) تصميم صفحات تُدوّن فيها الاستجابات: ينبغي أن يظهر كل مفهوم في صفحة مستقلة، حيث يُكتب المفهوم أعلى الصفحة يليه الموازين المتعلقة به، بحيث يكون مقابلاً لمنتصف كل ميزان، كما هو موضح بشكل (١٢ ـ ٣) .كما ينبغي أن يحدث تبادل عشوائي بين أقطاب الموازين أي بين أزواج الصفات مثل (موجب ـ سالب، غير ناضج ـ ناضج) ، وأن يكون عدد النقاط الفاصلة بين كل صفتين ثابتًا، ويفضل أن يكون سبع نقاط أو ربما تسع نقاط. غير أن هذا يعتمد على خصائص عينة الأفراد، فإذا كانوا أطفالاً صغارًا ربما يكتفى بخمس نقاط، ويمكن الاكتفاء بأزواج من الصفات عددها يتراوح بين ١٠ ، ١٥ زوجا تتعلق بعدد من المفاهيم وليكن عشرة تقريبًا، ولكن يمكن أن يستجيب الأفراد لعدد أكبر من المفاهيم وليكن ٢٠ مفهومًا بموازين عددها ٢٠ ميزانًا، ويفضل تقليل هذا العدد بالنسبة للأطفال.

(٤) كتابة تعليمات الاستجابة : ينبغي أن تشتمل صفحة الغلاف على تعليمات عامة تتعلق بالهدف من المقياس والمطلوب من المستجيب، وكذلك كيفية تدوين الاستجابات، وتوضيح ضرورة التعبير عن الانطباع الأول بسرعة، وإصدار أحكام مستقلة لكل زوج من الصفات المتعلق بالمفهوم المعين، والتعامل مع كل مفهوم بطريقة مستقلة، والتنويه بأن الفرد ينبغي عليه الحكم على المفهوم كما يراه وليس كما يراه أو يستجيب له الآخرون، مع مراعاة عدم تفسير العلاقة بين الصفة والمفهوم نظرًا لأن هذا يؤدي إلى عدم صدق استجابة الفرد. فمثلاً يمكن أن تكون التعليمات كالتالي:

الهدف هو قياس معانى بعض المفاهيم، والمطلوب منك الحكم عليها استنادا إلى سلسلة من الموازين الوصفية. وسوف تجد فى كل صفحة مفهوما مختلفا يليه مجموعة من الموازين، وعليك أن تُقدّر كل مفهوم وتضع علامة (X) فوق الخط الذى يناسب هذا التقدير كالتالى:

خامل :___ :___ :___ :___ :___ :_X_ :___ نشط

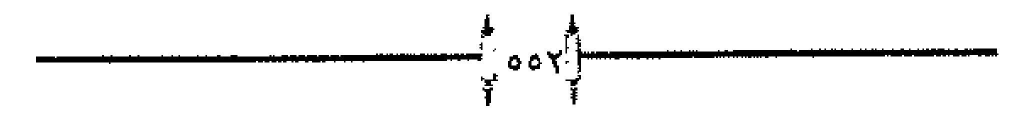
قوى :___ :_X_ :___ :___ :___ :___ :___ ضعيف

(٥) تقدير درجات الموازين والمفاهيم : لتقدير درجات الاستجابات تُعيّن قيم تتراوح بين ١ ، ٧ إذا كان عدد نقاط كل ميزان سبع نقاط، أو بين ١ ، ٩ إذا كان عددها تسع نقاط وهكذا، بحيث إن النقطة القريبة من الصفة التى تُمثل القطب السالب (تقييم منخفض) يُعيّن لها الدرجة (١) ، والنقطة القريبة من الصفة التى تُمثل القطب الموجب (تقييم مرتفع) يُعيّن لها الدرجة (٧)، وتُحدد درجات الفرد وفقًا لمواقع العلامات التى وضعها. فإذا كان لدينا ١٠ موازين لمفهوم معين، فإن أقصى درجة تكون ٧٠ ، وأقل درجة ١٠.

ويمكن استخدام هذه الدرجات فى المقارنة بين اتجاهات الفرد نحو مفاهيم مختلفة مثل مفهوم محاضرة رياضيات، ومفهوم فيلم علمى عن الكيمياء. كما يمكن مقارنة تقديرات فردين مختلفين لمفهوم معين، وكذلك يمكن ترتيب درجات مجموعة من الطلاب لتقييم الفروق بينهم فى الاتجاه، ويمكن أيضًا عن طريق إعادة تطبيق الموازين على المجموعة بعد مرور مدة زمنية معينة تقييم التغيرات التى حدثت فى اتجاههم.

مزايا وعيوب أسلوب تمايز معانى المفاهيم :

يُعد أسلوب تمايز معانى المفاهيم من الأساليب المفيدة فى تقييم ما يراه الفرد أكثر جاذبية من المفاهيم والمواقف الاجتماعية والشخصية. والاهتمام بعامل "التقييم" فى استجابات الأفراد على الموازين المختلفة يفيد فى تعرُّف رد فعل الفرد نحو كثير من الأشياء، إذ يمكن أن يُطلب منه تقدير أفراد عائلته، وزملائه فى العمل، وأصدقائه ومعلميه، وأصحاب المهن الأخرى، والأنشطة، والأفكار المجردة، مثل الانتماء، والسلام، والديمقراطية، وكذلك البرامج المرئية، والمسموعة، والمقروءة، وغير ذلك مما يصعب تقييمه باستخدام الأساليب الأخرى التى تتطلب إجراءاتها جهدًا كبيرًا. غير أن هذا يتطلب العناية باختيار الموازين التقييمية لكى لا يؤدى ذلك إلى أخطاء فى التقدير. فمثلا ينبغى التحقق من أن الصفتين المتناقضتين يمثلان قُطْبى الميزان ويقعان



بالفعل على متصل متدرج، فأحيانًا توجد صفات لها أكثر من نقيض، وصفات يصعب اختيار نقيضها أو ربما ليس لها نقيض. وقد أوضحت كثير من الدراسات أن صدق موازين تمايز معاني المفاهيم وثبات درجاتها مُرضيًا بوجه عام، والارتباطات بين التقديرات المتعلقة بهذه الموازين ودرجات موازين ثيرستون وليكرت تراوحت قيمها بين ٧٤,٠ ، ٨٢,٠ ، كما أن قيم معامل الثبات باستخدام طريقة التجزئة النصفية تراوحت بين ٨٣,٠ ، ٩١,٠ .

وعلى الرغم من ذلك فإن هناك بعض التساؤلات حول هذا الأسلوب يتعلق بعضها بما إذا كان يفيد بالفعل فى قياس المعنى، فقد وُجد أن بعض المفاهيم تنال تقديرات متماثلة من معظم الأفراد على الرغم من أنها ربما لا تكون متماثلة تماثلاً تامًا فى المعنى وذلك لأن التقديرات تعتمد على معنى المضمون Connotative Meaning ، وليس على المعنى بجميع جوانبه. كما أن بعض الصفات الثنائية ربما لا تعبر تعبيراً معقولاً عن مفهوم معين، فمثلا إذا كان المفهوم «الديمقراطية» وكان الميزان ثنائى القطب «عادل ـ غير عادل» أو «أمين ـ غير أمين» فإن الأساس الذى يستند إليه تقدير الأفراد يكون غير واضح.

أسلوب التصنيف الترتيبى The Q Sort :

يُستخدم هذا الأسلوب أيضًا فى قياس الاتجاهات والميول وغيرها من الجوانب الوجدانية، وقد اقترحه ستيفنسون (Stephenson, 1953 , 1967) . ويفيد بدرجة أساسية فى أغراض الإرشاد والعلاج النفسى، والتغيرات السلوكية، والمقارنة بين تفضيلات أو آراء مجموعات مختلفة من الأفراد فيما يتعلق بموضوع ما، إلا أنه يختلف فى منهجيته عن أسلوب تمايز معانى المفاهيم، فهذا الأسلوب يهتم بالمقارنة ودراسة التماثل بين التقديرات المتعلقة بالاتجاهات والمشاعر بين الأفراد المختلفين. ويعتمد هذا الأسلوب على ترتيب مجموعة من الفقرات أو العبارات، أو الأشياء، أو الأشخاص أو غير ذلك وفقًا لمحك معين بواسطة الأفراد. وفى حالة قياس الاتجاهات يكون المحك هو شدة الموافقة، أو الأهمية، أو الأفضلية. وما يُميِّز هذا الأسلوب هو أن التراتيب التى نحصل عليها بواسطة الأفراد ينبغى إيجاد الارتباط فيما بينها والتوصل إلى تجمعات من الأفراد الذين تكون تراتيبهم متماثلة للكشف عن العامل الذى يمكن أن يفسر أنماط هذه التراتيب، وبذلك يختلف عن غيره من الأساليب فى أنه يعتمد على الارتباطات بين الأفراد وليس على الارتباطات بين المقاييس.

ويتطلب هذا الأسلوب إعطاء الفرد مجموعة من البطاقات مُدوَّن على كل منها فقرة أو عبارة، أو سمة أو صورة أو غير ذلك، ويُطلب منه تصنيف هذه البطاقات وترتيبها فى مجموعات وفقًا لموقعها النسبى على مُتصل أو بُعد واحد يتراوح بين «أكثر أهمية» و «أقل أهمية» ، أو بين «أكثر مشابهة» و «أقل مشابهة»، أو «أكثر تميزًا» و«أقل تميزًا» أو «أفضل» ، و «أسوأ» وهكذا. ويتم تحديد عدد البطاقات التى يسمح بها فى كل مجموعة مسبقًا بحيث يقترب توزيعها من التوزيع الاعتدالى. ولعل هذا يماثل الخطوة الأولى فى الإجراءات المتعلقة بأسلوب ثيرستون حيث يقوم المحكمون بتصنيف الفقرات فى ٩ أو ١١ مجموعة بهدف تحديد قيم ميزان كل فقرة استنادًا إلى هذا التصنيف، أما فى هذا الأسلوب فإن الهدف من التصنيف معرفة وجهة نظر الفرد أو اتجاهه نحو الشىء أو الموضوع المعين.

ونظرًا لأن هذا الأسلوب يركز فقط على بُعد أحادى، والصعوبات المتعلقة ببناء المقاييس وتطبيقها وتحليل بياناتها استنادًا إليه يقلل من استخدامه فى أغراض التقييم إذا ما قورن بغيره من الأساليب الشائعة الاستخدام.

أسلوب البسط Unfolding Technique :

اقترح كومبس (Coombs, 1953 , 1964) أسلوبًا جديدًا فى إطار نظريته التى أطلق عليه «نظرية البيانات Theory of Data» يُعرف باسم أسلوب البَسْط Unfolding Technique. ويُستخدم فى استخلاص سمة كامنة مثل الاتجاهات تنطوى عليها تفضيلات مجموعة من الأفراد. ويعتمد هذا الأسلوب على تجميع عدد من الفقرات أو العبارات التى يُفترض أنها تتعلق بالاتجاه المراد قياسه. ويُطلب من الأفراد اختيار الفقرات التى يرون أنها أكثر تفضيلاً بالنسبة لهم. ويُفترض أن كل فرد يختار الفقرات التى تقترب من موقعه على متصل الاتجاه. ونظرًا لأننا نتغاضى عن توجُّه الاتجاه، فكأن الفرد طوى Fold ميزان الاتجاه Attitude Scale عند النقطة التى تحدد موقعه على المتصل لكل من الأفراد والفقرات. ويوضح شكل (١٢ـ٤) التالى مثالاً لهذا الميزان:



شكل (١٢ـ٤) يوضح مثالا افتراضيًا للميزان (أ) لمواقع ثلاث فقرات وثلاثة أفراد على متصل اتجاه معين

ويوضح هذا الشكل توزيعًا افتراضيًا لثلاث فقرات (أ) ، (ب) ، (جـ) استجاب لها ثلاثة أفراد (١) ، (٢) ، (٣) على متصل اتجاه معين، وتمثل الأسهم نقط المنتصف بين أزواج الفقرات. فمثلاً إذا طُلب من الفرد (١) أن يختار الفقرتين الأفضل بالنسبة له، فإنه سوف يختار الفقرتين (أ) ، (ب) على افتراض أن هذا الاختيار يعتمد على المسافة النسبية بين موقعه على المتصل وموقع الفقرتين اللتين اختارهما، كذلك الفرد (٣) سوف يختار الفقرتين (ب) ، (جـ) . أما الفرد (٢) فعلى الرغم من أن موقعه على المتصل إلى يسار كل من الفقرتين (أ) ، (ب)، إلا أنه سوف يختار أيضًا الفقرتين (أ) ، (ب)، وهذا يصبح واضحًا إذا تصورنا أننا قمنا بطىّ المتصل عند موقع الفرد (٢) عندئذ تكون الفقرتان (أ) ، (ب) أقرب إلى موقعه من الفقرة (جـ) بسبب المسافات النسبية بين الفقرات الثلاث.

وإذا نظرنا إلى الشكل سوف نرى أن جميع الأفراد إلى يمين نقطة المنتصف أجـ سوف يختارون الفقرتين (أ) ، (ب) ، بينما الأفراد إلى يمين نقطة المنتصف ب جـ ولكن إلى يسار نقطة المنتصف أجـ سوف يختارون الفقرتين (ب) ، (جـ) ، وبذلك يمكن استخدام توزيع اختيارات الأفراد لتحديد المسافات النسبية بين الفقرات على متصل الاتجاه.

وعلى الرغم من أن هذا الأسلوب يهدف لكشف السمات الكامنة، فإنه يمكن استخدامه فى بناء موازين رتبية للاتجاهات استرشاداً بالخطوات السابقة، ومن الجدير بالذكر أن استخدام هذا الأسلوب والإفادة منه يتطلب الرجوع إلى كتاب نظرية البيانات الذى أشرنا إليه.

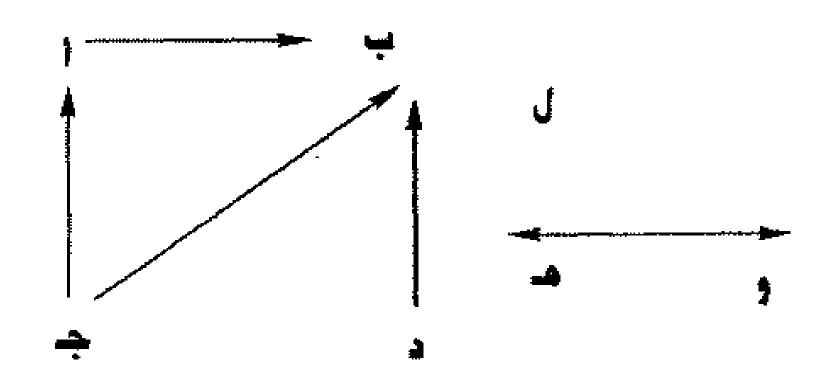
أساليب قياس العلاقات الاجتماعية Sociometric Techniques :

تهدف هذه الأساليب لدراسة التفاعل الاجتماعي بين أفراد مجموعة معينة، وقياس أنماط هذا التفاعل والتواصل. ويمكن عن طريق هذه الأساليب جمع معلومات عن توجهات التفاعل بين أعضاء مجموعة واحدة أو مجموعات فرعية، أو بينها وبين كل فرد على حدة، ويمكن أن يكون التفاعل سلوكيًا فعليًا أو ما هو مرغوب أو متوقع أو متصور.

ونوع هذا التفاعل أو محتواه يتخذ أشكالا مختلفة من السلوك الاجتماعى، مثل الجلوس بجوار، تناول الطعام مع، الشراء من، الإقراض، الزيارة، اللعب مع، الصداقة مع، التحدث مع، السكن بجوار، وغير ذلك. ويمكن الإفادة من ذلك فى إطار العمل التربوى، حيث إن البنية الاجتماعية للصف الدراسى تتكون من أنماط

مخـتلفة من العـلاقات التي تتطلـب الدراسة، فالطلاب الـذين يشعرون بالارتـياح مع أقرانهم يستطيعون أن يستثمروا قدراتهم بـدرجة أفضل ممن يشعرون بالضيق منهم. كما أن سلوك المعـلم وعلاقاته مع طلابه يؤثر فـي إدراك الطلاب لمعلمهم وفي تـعلمهم. واتجاهات الـطلاب نحو مدرسـتهم يكون في كثـير من الأحيـان انعكاسًا لمـدى تقبل أقرانهم ومعلمـيهم لهم. لذلك فإن البيانـات التي نحصل عليها من مقـاييس العلاقات الاجتماعية تفيد في دراسة الأنماط التنظيمية المثلى لتكوين مجموعات فاعلة، والتكيف الاجتماعي للـطلاب، والبنية الاجتماعية للـمجموعات، وتأثير المدرسـة والممارسات الصفية والتجديدات التربوية في العلاقات الاجتماعية داخل المدرسة، والعلاقات القائمة بين مكانة الطلاب داخـل جماعتهم،والقدرات العـقلية والتحصيل الدراسـي،وتفضيلات المعلمين لطلابهم. وتـتعدد أساليب قياس العلاقات الاجتماعيـة حيث يمكن استخدام أساليب المـلاحظة الفعلية للأفراد أثنـاء قيامهم بأنشطتهـم المعتادة لتعرف طبـيعة هذه الأنشطة والعلاقات بين الأفراد المشاركين فيها من حيث المبادأة والتواصل. كما يمكن استخدام استبيانات التفاعل الاجتماعي Sociometric Questionnaires ، والمقابلات الشخصية Interviews ، حيث يُطلب من كل فرد في المجموعة أن يحدد الأفراد الذين يفضل مشاركتـهم له في بعض الأنشطة، والأفراد الذين لا يفضل مشـاركتهم، وأحيانًا يطلب مـن الفرد تحـديد أي عدد من الأفـراد، ولكن عادة يكـون عدد الأفراد مـحدد مسبقًا، ويُفـترض في مثل هذه الحالات أن يكون الفرد أمينًا في تـفضيلاته؛ لأنه سوف يترتب على ذلك إجراء تنظيمات معينة تتعلق بالجماعة التي ينتمي إليها الفرد.

ولعل الأسلوب الذي اقترحه مورينو ومعاونوه Moreno et al. يُعد من الأساليب المبكرة التي تتعلق بالتمثيـل الشكلي للعلاقات الاجتماعية Sociograms، التي تُعرف في وقتنا الحاضر بـاسم الأشكال الموجهة Directed Graphs ، ويستخدم في تحليل بنية العلاقات بين أفراد جماعة معينة .ويمكن توضيح ذلك بالشكل (١٢ـ٥) التالي:



شكل (١٢ـ٥) يوضح العلاقات بين أفراد جماعة معينة

ويتضح من شكل (١٢ـ٥) أن الجماعة تشتمل على سبعة أفراد تربط بينهم علاقات ممثلة بالأسهم، وبعضهم لا تربطه علاقات حيث لا توجد أسهم متجهة منه أو إليه. فالفرد (أ) تربطه علاقة بالفرد (ب)، والفرد (جـ) تربطه علاقة بكل من (أ) ، (ب)، والفرد (هـ) تربطه علاقة متبادلة مع الفرد (و)، والفرد (ل) لا تربطه علاقة بأحد، وهكذا.

وتستخدم مثل هذه الأشكال إذا كانت الجماعة قليلة العدد، أما إذا كانت تشتمل على عدد كبير نسبيًا (١٠ أفراد أو أكثر) وبنية العلاقات الاجتماعية أكثر تعقيدًا، فإنه يفضل في هذه الحالة استخدام مصفوفات العلاقات الاجتماعية Sociometic Matrices. ويحدد على كل من بعدى المصفوفة الأفقى والرأسى الحروف المناظرة للأفراد، ويدون فى خلاياها الرقم (١) إذا كانت هناك علاقة بين الفرد وغيره من الأفراد، والرقم (صفر) إذا لم تكن هناك علاقة بينه وبين أى منهم، ويجرى تحليل بيانات المصفوفة من حيث العلاقات الأحادية والثنائية ، وعدم وجود علاقات.

واقترح مورينو مؤشرًا لجاذبية الفرد فى جماعته وذلك بقسمة العدد الكلى للاختيارات والرفض، وقسمة الناتج على العدد الكلى للاختيارات، كما اقترح مؤشرًا لتماسك الجماعة Group Cohesiveness Index وذلك بقسمة عدد الاختيارات المتبادلة مثنى مثنى على عدد الأزواج الممكنة [$\frac{٥(٥-١)}{٢}$] .

وفيما يتعلق بصدق مقاييس العلاقات الاجتماعية ، فإنه ربما يصعب التحقق منها إلى حد ما. أما الثبات فقد أوضحت بعض الدراسات أن اتساق الاختيارات بمرور الزمن يبدو مرتفعًا إلى حد ما، غير أن هذا يتأثر بالمستوى العمرى لأفراد الجماعة، واستقرارها ومدى شدة التفضيل.

أساليب بناء موازين الاتجاهات متعددة الأبعاد :

تُعدُّ جميع أساليب بناء موازين الاتجاهات التى ناقشناها فيما سبق أساليب أحادية البعد، أى تحاول تحديد موقع الفرد على متصل ذو بُعد تقييمى واحد. وعلى الرغم من أن هذا يتفق مع تعريف الاتجاه الذى قدمناه فى مستهل هذا الفصل، إلا أن بعض علماء النفس عرّفوا الاتجاه بأنه مفهوم مركب متعدد الأبعاد، ويتطلب قياس الاتجاه بهذا المفهوم أساليب متعددة الأبعاد لبناء موازين الاتجاهات Multidimensional Scaling Procedures.

وتهدف هذه الأساليب لتحديد الأبعاد المناسبة التي تنطوي عليها استجابات الأفراد فيما يتعلق بموضوع الاتجاه أو مجموعة من الموضوعات. وعلى الرغم من أننا نتوقع أن هذه الأبعاد المتعددة تختلف باختلاف الأفراد والموضوعات، إلا أنها لا يجب أن تتغير بتغير مجموعة معينة من الاستجابات.

ويُعدّ أسلوب التحليل العاملي Factor Analysis من الأساليب المتعارف عليها في استخلاص الأبعاد المتعددة من مجموعة من البيانات، وقد استخدمه أوزجود Osgood كما سبق أن أشرنا في استخلاص أبعاد تمايز معاني المفاهيم. وقد أوضحنا في كثير من الفصول السابقة استخدامات هذا الأسلوب في مجال القدرات العقلية والاستعدادات الخاصة والميول. كما يُعد أسلوب الموازين متعددة الأبعاد Multidimensional Scaling من الأساليب الأخرى المماثلة للتحليل العاملي، ولكنه يعتمد على المسافات بين مجموعة من النقاط التي تُمثل الدرجات، فهو أسلوب إحصائي لا بارامتري مناظر لأسلوب التحليل العاملي الذي يعد من الأساليب الإحصائية البارامترية، ويهدف أيضًا للتوصل إلى مجموعة من الأبعاد. وموقع كل متغير يمثل الاتجاه على هذه الأبعاد، وبذلك يمكن تحديد عدد من الأبعاد التي تنطوي عليها معتقدات الفرد التي يُستند إليها في الحكم على موضوعات اتجاهات مختلفة. ومن الجدير بالذكر أن هناك برامج حاسوب يمكن استخدامها في إجراء التحليل متعدد الأبعاد يتعلق بهذا الأسلوب، مثل برنامج INDSCAL الذي أعده كارول، وشانج (Carroll & Chang , 1970) . وقد أشرنا في الفصل الخامس إلى البرنامج الذي يجري التحليل العاملي.

تعقيب عام على أساليب بناء موازين الاتجاهات :

لعلنا لاحظنا مما سبق تنوّع أساليب بناء موازين الاتجاهات وتباين منهجياتها وإجراءاتها حيث إن كلا منها يستند إلى إطار نظري معين، غير أن معظمها يؤدي إلى درجات للاتجاه تمثل مواقع الأفراد على متصل تقييمي ثنائي القطب يتعلق بموضوع الاتجاه. ومع هذا فإنها تختلف فيما بينها من حيث مستويات القياس، وخصائص الفقرات التي يتم اختيارها، وكذلك الخصائص السيكومترية.

فقد اعتمد أسلوب ثيرستون على أحكام مجموعة من الخبراء في تقدير قيم ميزان الفقرات التي تُمثل موقع كل مفردة على متصل الاتجاه، وافترض أن هذه الأحكام تكون على ميزان متساوي الفترات مما يترتب عليه افتراض أن قيم ميزان الفقرات ودرجات الاتجاه تكون فتراتها متساوية، في حين أن الأسلوب التراكمي الذي اقترحه جتمان يؤدي إلى ترتيب الفقرات على متصل الاتجاه دون افتراض تساوي الفترات،

لذلك فإن ميزان الاتجاه يكون من المستوى الرتبي، أما أسلوب التقدير الجمعي الذى اقترحه ليكرت فإنه لا يعكس درجات متباينة من الأفضلية أو الموافقة، وإنما يصنف الاستجابات إلى استجابات موجبة أو سالبة، ومع هذا يؤدى إلى ميزان رتبى أيضاً. وفيما يتعلق بأسلوب تمايز معانى المفاهيم الذى اقترحه أوزجود، فإنه يعتمد على فقرات (صفات) ثنائية القطب، ولذلك لا نستطيع تصنيفها إلى فقرات موجبة وأخرى سالبة، وإنما يُفترض فقط أنها ترتبط بالبُعد التقييمى للفضاء السيمانتى Semantic Space الذى يشتمل على ثلاثة أبعاد كما سبق أن أشرنا. أما النقاط أو المواقع السبعة التى تقع على الميزان ثنائى القطب، فإنه يُفترض أنها تُمثل فترات متساوية وبالتالى يُفترض أن درجات الاتجاه تفصلها فترات متساوية كما فى أسلوب ثيرستون.

وهذا الاختلاف فى مستويات القياس يؤدى إلى اختلاف فى الإفادة من الفقرات المحايدة Neutral Items فى تقييم الاتجاهات. ففى أسلوب ثيرستون تُعد الفقرات المحايدة مؤشرات صادقة للاتجاه لأنها تُشكل جزءاً من ميزان الاتجاه، كذلك يمكن أن يشتمل ميزان الاتجاه وفقاً لأسلوب جتمان على فقرات محايدة، ولكننا لا نستطيع تمييزها عن غيرها من الفقرات نظراً لأن أسلوب جتمان يهتم بالتحقق من إمكانية ترتيب الفقرات على بُعد أحادى، لذلك يمكن أن تكون هناك موافقة أو عدم موافقة على جميع الفقرات أو تتوزع بين الموافقة والمحايدة وعدم الموافقة، أما الفقرات التى لا نستطيع تصنيفها إلى فقرات موجبة أو سالبة وفقاً لأسلوب ليكرت فإنه يتم حذفها، حيث يفترض أن الفقرات المحايدة لا تشير بصدق إلى الاتجاه. كذلك فى أسلوب تمايز معانى المفاهيم تُستبعد الصفات ثنائية القطب (الفقرات) التى لا تنتمى بوضوح إلى بُعد «التقييم» ومع هذا يسمح بالاستجابة المحايدة على هذا البُعد. وجميع الأساليب تهدف لتحديد الفقرات الواضحة المتعلقة بالاتجاه المراد قياسه، ولكن اختلاف محكات اختيار الفقرات يؤدى إلى خصائص مختلفة للمنحنيات المميزة Trace lines لهذه الفقرات. ففى كل من أسلوب جتمان وليكرت ، وأسلوب تمايز معانى المفاهيم تكون هذه المنحنيات المميزة اطرادية Monotonic، أى أنه كلما زادت درجة الاتجاه يزداد احتمال موافقة الفرد على الفقرة. بينما فى أسلوب ثيرستون تكون هذه المنحنيات غير إطرادية Nonmonotonic،أى أن احتمال الموافقة على فقرة معينة يُفترض أن يزداد كلما اقتربت درجة الاتجاه للفرد من قيمة ميزان الفقرة Item's Scale Value من أىٍّ من طرفى المتصل ثنائى القطب، مما يجعل المنحنى المميز للفقرة يتخذ شكل حرف U.

كذلك فى أسلوب جتمان تكون الفقرات متراكمة، بينما فى الأساليب الأخرى لا تكون كذلك، ولهذا فإن كل درجة من درجات الاتجاه فى الميزان التراكمى لجتمان

تقترن بنمط واحد فقط من الاستجابات للفقرات على ميزان الاتجاه، بينما يمكن الحصول على أنماط متعددة للاستجابات للدرجة نفسها في الأساليب الأخرى، لذلك لا نستطيع التنبؤ بالاستجابات للفقرات من درجات الاتجاه إلا في حالة استخدام أسلوب جتمان.

ولعل طبيعة موازين الاتجاهات التي تستند إلى جميع هذه الأساليب يمكن أن تضمن درجة عالية من الموضوعية، فجميع الاستجابات يمكن تقدير درجاتها باستخدام مفاتيح معينة أو باستخدام برامج حاسوب.

وسوف نلخص الاختلافات التي ذكرناها بين الأساليب الأربعة الأساسية في بناء موازين الاتجاهات في جدول (١٢ـ٤) التالي:

	الأسلوب			
	ثيرستون	جتمان	ليكرت	تمايز معاني المفاهيم
الصيغة	فقرات	فقرات	فقرات	فقرات
الموقع على المتصل	نقاط على مدى المتصل	نقاط على مدى المتصل	طرفي المتصل فقط	-
مستوى تكميم الفقرات	فترى (اعتبارى)	رتبى	نوعى	-
مستوى تكميم الدرجات	فترى (اعتبارى)	رتبى	رتبى	فترى
طبيعة المتصل	غير تجمعى	تجمعى	غير تجمعى	غير تجمعى
وجود فقرات محايدة	نعم	نعم	لا	لا
المنحنى المميز	غير اطرادى على شكل حرف U	اطرادى درجى	اطرادى خطى	اطرادى خطى

جدول (١٢ـ٤) يوضح جوانب الاختلاف بين الأساليب الأساسية في بناء موازين الاتجاهات

ثبات وصدق موازين الاتجاهات :

فيما يتعلق بثبات درجات موازين الاتجاهات، يُعد معامل الاتساق الداخلي، ومعامل الاستقرار من المعاملات المناسبة لتقدير ثبات درجات موازين الاتجاهات. فمعامل الاتساق الداخلي لدرجات ميزان اتجاه يشتمل على ١٠ فقرات واعتُنى بإعداده يمكن أن تقترب قيمته من ٠,٨٠ (Anderson, 1981)، وإذا زاد عدد الفقرات إلى ٢٠ فقرة يمكن أن تصل قيمة هذا المعامل إلى ٠,٩٠ ، لذلك يمكن بناء موازين اتجاهات تتميز درجاتها باتساق داخلي معقول. غير أن القيم التقديرية لمعامل استقرار الدرجات يُستخدم بدرجة أقل، وربما يرجع سبب ذلك إلى مشكلة تفسير هذه القيم. فمن الضرورى قبل تفسيرها معرفة درجة الاستقرار المتوقعة للاتجاه كتكوين فرضى عبر المدة الزمنية التى تفصل بين مرتى تطبيق الميزان. فالأدلة البحثية القليلة المتوافرة توضح أنه يمكن الحصول على قيم معقولة عبر مدة زمنية تتراوح بين أربعة وخمسة أسابيع (Anderson, 1981). أما أسلوب جتمان فإنه يتضمن داخله تقديرات الثبات، حيث يتطلب حساب معاملين أحدهما معامل إمكانية إعادة التكوين Reproducibility Coefficient ، والآخر معامل إمكانية بناء الميزان تراكميًا Scalebility Coefficient ولتقدير هذين المعاملين معًا يتم التحقق من توافر خاصية التراكم التى سبق أن أشرنا إليها، وهذان المعاملان يتعلقان بالاتساق الداخلى لميزان الاتجاه.

غير أن المشكلات الرئيسة لموازين الاتجاهات تتعلق بصدق هذه الموازين، فقد سبق أن أوضحنا فى الفصل الحادى عشر أن مشكلات الأنماط العقلية للاستجابات مثل مراعاة القبول الاجتماعى تشير إلى الاستجابات التى لا تتسق مع المشاعر الحقيقية للفرد، ولكن الفرد يلجأ إلى تزييف استجاباته بما يراه متوقعًا أو مقبولاً، وكذلك الإذعان أو الموافقة على معظم فقرات ميزان الاتجاه يلجأ إليها الفرد إذا لم يكن متأكدًا أو لديه مشاعر متناقضة.

وعلى الرغم من أن هذه الأنماط العقلية للاستجابات تُقلل من صدق موازين الاتجاهات، إلا أنه يمكن باستخدام فروض مناسبة تتعلق بهذه الأنماط التحقق الإمبيريقى من صدق التكوين الفرضى لهذه الموازين. ويمكن كذلك التحقق بسهولة من صدق محتوى الموازين التى تستند فى بنائها إلى كل من أسلوبى ثيرستون وليكرت، إذ يمكن إعادة كتابة الفقرات ومراجعتها عدة مرات للتوصل إلى اتفاق المحكمين على وضوحها وملاءمتها . كما يمكن باستخدام تحليل المفردات فى أسلوب ليكرت، وإيجاد قيمة نصف المدى الربيعى فى أسلوب ثيرستون الكشف عن الفقرات الغامضة.

كذلك يمكن التحقق من الصدق التلازمي لموازين الاتجاهات عن طريق تحديد مدى تمييز ميزان الاتجاه بين مجموعتين من الأفراد يكون معلومًا اختلافهم في الاتجاه المراد قياسه. وكذلك يمكن التحقق من ذلك بإيجاد الارتباط بين درجات ميزان اتجاه معين وموازين أخرى تتعلق بموضوعات تماثل موضوع الاتجاه المراد قياسه، مثل ميزان اتجاه نحو التدين، وميزان اتجاه نحو أماكن العبادة.

إرشادات عامة لبناء موازين الاتجاهات:

اتضح لنا مما سبق أن موازين الاتجاهات تشتمل على مجموعات من الفقرات أو العبارات يستجيب لها الفرد على ميزان معين، وهذه الفقرات أو العبارات تتعلق بموضوع أو شخص أو خبرة أو قضية معينة، وتعتمد نوعية الميزان على نوعية فقراته. لذلك ينبغي العناية ببناء وصياغة هذه الفقرات استرشاداً بالمقترحات أو المحكات التالية التي ينطبق بعضها على جميع موازين الاتجاهات، بينما ينطبق البعض الآخر على نوع معين من الموازين (Babbie, 1973 , Edwards, 1957):

(١) تجنب الفقرات أو العبارات التي تشير إلى حقائق، حيث إن الفقرات المتعلقة بالاتجاهات تتطلب أن يعكس الفرد مشاعره ومعتقداته حولها وليس تقرير صحة أو خطأ الفقرات، فتقرير ذلك يتعلق بمقاييس الجوانب المعرفية بينما يجب تجنب ذلك في المقاييس الوجدانية.

(٢) تجنب الفقرات التي يكون لها تفسيرات متعددة، حيث إن الفقرة ينبغي أن تحتمل تفسير واحد فقط لكي يعرف الفرد كيف يستجيب لها.

(٣) تجنب الفقرات غير المتعلقة بموضوع الاتجاه المراد قياسه.

(٤) ينبغي أن تكون جميع الفقرات واضحة ومصاغة بلغة سلسة وبسيطة تناسب أعمار المختبرين بحيث يمكنهم فهمها، حيث إن هذه الفقرات لا تهدف لقياس الذكاء أو الفهم.

(٥) ينبغي أن تكون كل فقرة أحادية البُعد، أي تتعلق بمفهوم واحد فقط، كما ينبغي أن لا تكون مزدوجة مثل «لا أحب المعلم الذي يعامل الطلاب معاملة غير عادلة ولا يسمح لهم بالمناقشة»، فالفرد ربما يوافق على أحد جزئي الفقرة، ولا يوافق على الجزء الآخر.

(٦) تجنب اقتراح استجابة معينة، مثل «هل توافق على أن التعليم المجاني يشجع الطلاب على التكاسل؟».

(٧) تجنب الفقرات التي يحتمل أن يوافق عليها أو لا يوافق عليها جميع الأفراد، فمثل هذه الفقرات لا تميز بين درجات الموافقة أو الأفضلية، لذلك ينبغي اختيار فقرات تغطي المدى الكلي لمتصل الأفضلية، فلا يجوز مثلاً أن تكون الفقرة «أحترم الأشخاص الذين يؤدون عملهم بأمانة».

(٨) ينبغي اختيار عدد متساوٍ من الفقرات الموجبة والسالبة بقدر الإمكان لتقليل تأثير تحيز الاستجابات.

(٩) تجنب استخدام كلمات تشير إلى عموميات مثل «دائمًا» ، «جميعًا» ، «أبدًا» «مطلقًا» حيث إن اشتمال الفقرات عليها يحدد استجابات الموافقة أو غير الموافقة آليًا.

(١٠) ينبغي الحيطة عند استخدام كلمات مثل «فقط» ، «بقدر الإمكان» ، «مجرد» ، أو التقليل من استخدامها.

(١١) ينبغي أن تكون الفقرات قصيرة، ويُفضل ألا يزيد عدد كلماتها عن ٢٠ كلمة.

(١٢) تجنب استخدام فقرات تشتمل على نفي مضاعف، أي نفي النفي، حيث إن مثل هذه الفقرات يصعب فهمها، والأفضل جعلها موجبة.

(١٣) ينبغي أن تشير الفقرات إلى الحاضر أو المستقبل وليس الماضي، وذلك لأن قياس الاتجاهات يتعلق بالوضع الراهن للاتجاهات وآثارها المتوقعة.

(١٤) ينبغي توزيع الفقرات في ميزان الاتجاه عشوائيا، ولكن ينبغي التحقق من عدم وجود أربع أو خمس فقرات موجبة أو سالبة متتالية.

(١٥) ينبغي مراعاة أن تكون الفقرات الأكثر حساسية تقع بالقرب من منتصف ميزان الاتجاه.

(١٦) يُفضل ألا يزيد عدد نقاط متصل الاتجاه المراد قياسه عن سبع نقاط.

أنواع أخرى من مقاييس الاتجاهات :

تعتمد موازين الاتجاهات التي ناقشناها فيما سبق على تصميم مواقف افتراضية يستجيب لها الفرد، ونستدل من هذه الاستجابات على اتجاه الفرد نحو موضوع الاتجاه، غير أن استجابات الفرد ربما لا تتفق مع سلوكه الفعلي المتعلق بالاتجاه، كما أن موازين الاتجاهات تُقصر ردود فعل الفرد أو استجاباته على مدى ضيق نسبيًا من المحتوى، ولا تمنحه الفرصة للتعبير عن اتجاهاته بأسلوبه الشخصي. لذلك فإنه يمكن

قياس الاتجاهات بأساليب الملاحظة المباشرة Observational Techniques، وأساليب المقابلة الشخصية وجهًا لوجه . وتُعد هذه الأساليب وغيرها من الأساليب السابقة من النوع غير التنكُّرى Undisguised Techniques، حيث يكون الفرد عادة مدركًا أنه يجرى تقييم اتجاهه، أى أن هدف القياس يكون واضحًا باستثناء أسلوب تمايز معانى المفاهيم الذى يُقدَّم أحيانًا للمستجيبين على أنه مقياس للمعنى.

ولكن يمكن قياس الاتجاهات بأساليب أخرى من النوع غير المباشر أى التنكُّرى Disguised Techniques، حيث يكون هدف القياس خافيًا على المستجيب، وتبرير ذلك أنه إذا كان هدف أداة القياس خافيًا أو غير واضح، فإن احتمال تزييف الفرد لاستجاباته يكون أقل مما لو كان الهدف واضحًا، وبخاصة إذا كان سيترتب على النتائج قرارات تؤثر على الفرد، وبذلك يمكن الحصول على معلومات صادقة. وفى مثل هذه الحالات ينقاد الفرد إلى الظن بأن استجاباته سوف يتم تفسيرها كمؤشرات لمتغير آخر يختلف عن المتغير الذي يجرى قياسه أو تقييمه.

وتوجد أنواع متعددة من هذه الأساليب غير المباشرة سوف نشير إلى بعض منها فيما يلى:

أسلوب اختيار الخطأ، وقد اقترحه هاموند Hammond's Error-Choice Technique. ويعتمد هذا الأسلوب على جعل الفرد يظن أنه يتخذ حكمًا موضوعيًا، أى أن هناك إجابة صحيحة. ومثال ذلك اختبارات المعلومات Information Tests التى تبدو كما لو كانت تقيس معلومات الفرد ولكنها فى واقع الأمر تقيس الاتجاهات. وتُستخدم فى هذه الاختبارات مفردات اختيار من متعدد تتناول مشكلات معينة تبدو كما لو كانت مشكلات موضوعية، ولكنها تهدف لتقييم معتقدات الفرد. ويستند هذا الأسلوب إلى افتراض أن الاتجاهات تؤدى إلى تحيز الاستجابات بطريقة متسقة، ويمكن استخدام هذه الاستجابات المتحيزة فى الاستدلال على اتجاه الفرد.

أسلوب القبول الظاهرى : وقد اقترحه كوك Cook's Plausibility Technique. ولا يعتمد هذا الأسلوب على استجابات الموافقة أو عدم الموافقة لمجموعة من الفقرات، وإنما يُطلب من الفرد الحكم على كل فقرة من حيث صدقها وفاعليتها، ومتضمناتها التقييمية، ومعقوليتها الظاهرية. فمثلا مهمة المحكمين فى أسلوب ثيرستون تقييم استحسان كل فقرة وتصنيفها فى إحدى المجموعات التى عددها ١١ مجموعة. ويمكن الحصول على قياس غير مباشر للاتجاه، إذا استخدمنا تقييم المحكمين فى الاستدلال على اتجاههم، ويستند هذا الأسلوب إلى افتراض أن اتجاه

الفرد سوف يؤدى إلى تحيز حكمه على التقبل الظاهرى للفقرة، أى أن الفقرة سوف تكون معقوليتها الظاهرية بالنسبة للأفراد الذين يوافقون على القضية التى تؤيدها أكثر من معقوليتها بالنسبة للأفراد الذين لا يوافقون عليها.

أسلوب تقدير استجابات الآخرين : ففى هذا الأسلوب Estimation of Others Responses يطلب من الفرد الاستجابة لأى سؤال كما لو كان فرد أو مجموعة من الأفراد الآخرين يجيبون عن هذا السؤال. ويستند هذا الأسلوب إلى افتراض أن الأفراد يُسقطون معتقداتهم واتجاهاتهم ومقاصدهم على الأفراد الآخرين عند استجابتهم «كما يعتقد الآخرون»، ولكنهم فى واقع الأمر يستجيبون بالطريقة التى يعتقدونها أنفسهم.

ولا يتطلب هذا الأسلوب بناء أدوات قياس معينة، وإنما يمكن أن يطلب من الفرد أن يوضِّح الطريقة التى يعتقد أن معظم الأفراد الآخرين سوف يستجيبون لأى فقرة من فقرات ميزان اتجاه مقنن. ونظراً لأن تقديرات اعتقاده عن استجابات الآخرين يُفترض أنها مكافئة لاستجاباته عن نفسه، لذلك يمكن استخدام الطرق المعتادة لحساب درجاته فى ميزان الاتجاه.

أساليب تعتمد على اختبارات الإدراك والذاكرة: إذا كان الإدراك Perception انتقائيًا وكذلك الذاكرة، فإن الأفراد الذين لديهم اتجاهات مختلفة يجب أن يستجيبوا استجابات مختلفة لمثيرات متنوعة. وأحد هذه الأساليب يتطلب عرضًا سريعًا لصورة تفصيلية تشتمل على عناصر للاتجاه المراد قياسه، فمثلاً إذا أردنا دراسة الاتجاه نحو تحرُّر المرأة، فإنه يمكن أن نعرض على مجموعة من الأفراد صورة رجال ونساء يعملون فى أحد المكاتب الحكومية لمدة ثانيتين أو ثلاث ثوانى، ثم يطلب منهم إجابة بعض الأسئلة المفتوحة أو المقيدة المتعلقة بالأنشطة والأدوار التى يقوم بها كل من الرجال والنساء، كما يمكن أن تتطلب الأسئلة إجابات لا تتعلق مباشرة بالصورة نفسها (مثل: من الجالس بجوار المكتب؟ حيث لا يوجد مكتب بالصورة). غير أن تفسير الاستجابات التى نحصل عليها باستخدام هذا الأسلوب يعد أكثر صعوبة من الأساليب المباشرة أو غير التنكُّرية.

ومن الجدير بالذكر أن هناك أساليب إسقاطية متنوعة يمكن استخدامها فى قياس الاتجاهات وغيرها من السمات ودراستها، سوف نوضحها فى الفصل الثالث عشر عند مناقشتنا لأساليب قياس الشخصية الإنسانية.

تعقيب على الأساليب غير المباشرة في قياس الاتجاهات:

ينبغي الإشارة إلى أن هناك العديد من أوجه النقد التي تُوجَّه للأساليب غير المباشرة في قياس الاتجاهات وغيرها من السمات الإنسانية لعل أهمها هو الجانب الأخلاقي المتعلق بإخفاء هدف القياس عن المستجيبين (يمكن الرجوع إلى الفصل الثاني لمراجعة أخلاقيات القياس). كما أنه ليس هناك ما يُبرِّر استخدام هذه المقاييس المتخفية أو غير المباشرة في مجال الإرشاد أو العمل التربوي حيث تتطلب أن تكون العلاقة بين المرشد وعملائه أو المعلم وطلابه متميزة بالأمانة والانفتاح، واللجوء إلى إخفاء هدف المقياس لا يكون ضروريًا أو مطلوبًا من أجل الحفاظ على هذه العلاقة، علاوة على أن ابتكار أساليب أو مواقف غير مباشرة أو متخفية يتطلب جهدًا ووقتا كبيرين. كذلك أوضحت الدراسات صعوبة التحقق من صدق هذا النوع من المقاييس وفاعليتها في التمييز بين المجموعات المختلفة، والتنبؤ بالسلوك.

بعض مشكلات قياس الاتجاهات:

إن الاستناد إلى المنهجيات والأساليب والإرشادات التي أوضحناها فيما سبق يُعد من الأمور الأساسية في بناء موازين الاتجاهات، غير أنه مهما بلغت العناية بإعداد هذه الموازين تظل هناك عوامل دخيلة تؤثر في استجابات الأفراد وتقلل من صدق الميزان المتعلق بالاتجاه المراد قياسه. فالدرجة التي يحصل عليها الفرد في ميزان الاتجاهات Attitude Scale لا يمكن أن تكون خالية تمامًا من الخطأ، أو دقيقة الدقة الكاملة نظرًا لأنها تكون نتاج تفاعل معقد بين عوامل متعددة أحدها هو السمة أو الاتجاه المعين، ويتعلق بعض هذه العوامل بما سبق أن أشرنا إليه في الفصل الحادي عشر عند مناقشتنا لمشكلات قياس الميول، ونعني بذلك الاتجاهات العقلية في الاستجابات Response Sets. فهذه الاتجاهات العقلية تؤثر بطريقة منتظمة في استجابات الأفراد وتؤدي إلى تحيز هذه الاستجابات، أما العوامل الأخرى التي تؤثر في الاستجابات فقد سبق أن أوضحناها عند مناقشتنا لمفهوم الثبات في الفصل الرابع. لذلك سوف نقتصر فيما يلي على توضيح بعض الاتجاهات العقلية التي تؤثر في درجات الأساليب المباشرة في قياس الاتجاهات وبخاصة الموازين.

(١) النزعة إلى اختيار استجابة محايدة: فكثير من موازين الاتجاهات تشتمل على قسم «غير متأكد» أو «لا أدري» وينزع بعض الأفراد إلى اختيار هذا القسم ليعبِّر عن استجابتهم لجميع الفقرات. وقد وجد كرونباك Cronbach أن قيمة ثبات الدرجات التي تعتمد فقط على عدد الاستجابات المحايدة بلغت ٠,٧٣، لذلك فإن النزعة إلى اختيار هذا القسم المحايد Neutral Cotegory تُعد نزعة متسقة تؤدي إلى قدر كبير من التباين

الحقيقى فى هذه الدرجات. وربما يُفسَّر ذلك بأن بعض الأفراد يتوخون الحذر فى استجاباتهم، أو أنهم لا يستطيعون اتخاذ قرار، أو يفتقرون إلى الملاحظة، أو ربما يعنى أن الفرد ينزع إلى الاستجابة يمينًا أو يسارًا على متصل الاتجاه. ويُقدَّر أحيانًا وزن معين لمثل هذا القسم نظرًا لأن التحيز الناشئ عن الاستجابة المحايدة ربما يُعزى بدرجة كبيرة إلى عوامل مشتركة نهتم بقياسها، ولا يُعزى جميعه إلى الحذر أو أسباب أخرى لا تستند إلى أدلة. ويمكن التغلب على ذلك إلى حد ما بعدم استخدام هذا القسم المحايد فى ميزان الاتجاه لكى يضطر الفرد إلى الاستجابة إما إيجابًا أو سلبًا.

(٢) مراعاة القبول الاجتماعى: تتطلب جميع الأساليب المباشرة فى قياس الاتجاهات أن يقدِّم الفرد تقريرًا ذاتيًا عن اعتقاده أو اتجاهه نحو موضوع معين، وتوجد مواقف ربما يختلف اتجاهه أو اعتقاده نحوها عما هو سائد فى المجتمع، ففى مثل هذه المواقف ربما لا يلجأ الفرد إلى التعبير عن مشاعره الحقيقية، بل يستجيب بطريقة تتفق مع ما هو مقبول اجتماعيًا Socially Desirable. واختلاف إدراكات الأفراد فيما هو مقبول أو غير مقبول اجتماعيًا يمكن أن يؤدى إلى التقليل من صدق نتائج موازين الاتجاهات. وقد سبق أن أوضحنا عند مناقشتنا لمشكلات قياس الميول كيفية التغلب على هذا الاتجاه العقلى فى الاستجابة وغيره من الأنماط، ولعل من المناسب أيضًا أن نطلب من الأفراد الاستجابة دون ذكر أسمائهم لكى يعبِّروا بحرية عن مشاعرهم الحقيقية أو اتجاهاتهم إذا كان الهدف بحثيًا.

(٣) الاستجابات المتطرفة: تشير بعض الأدلة البحثية إلى وجود فروق متسقة بين الأفراد فى نزعتهم إلى استخدام الأقسام المتطرفة Extreme Categories فى موازين الاتجاهات، مثل «موافق جدًا» أو «غير موافق على الإطلاق». فبينما أشرنا أعلاه إلى نزعة بعض الأفراد إلى اختيار القسم الأوسط «محايد»، توجد نزعة لدى أفراد آخرين إلى اختيار الأقسام المتطرفة. وقد أجريت دراسات متعددة لإيجاد العلاقة بين هذه الاستجابات وبين متغيرات مثل الذكاء، والتحفظ، والوثوقية وغيرها من سمات الشخصية. ويبدو أنه من الصعب بناء ميزان اتجاهات لا يتأثر بهذه النزعة المحتمل حدوثها. غير أنه توجد بعض الأساليب الإحصائية التى يمكن استخدامها فى تحديد مدى تأثر استجابات الأفراد بها، غير أن النزعة للاستجابات المتطرفة لا تؤثر بعامة فى صدق الميزان إلا تأثيرًا طفيفًا.

(٤) الإذعان أو الموافقة: يشير هذا الاتجاه العقلى فى الاستجابة إلى نزعة الفرد إلى الموافقة على معظم الفقرات، وقد حظى هذا النمط بكثير من البحوث والدراسات، وبخاصة بعد أن اقترح أدورنو (Adorno, 1950) نظريته المتعلقة بتكوين بنية الشخصية السلطوية Authoritarian Personality التى أشرنا إليها فى مستهل هذا

الفصل، حيث تبيَّن أن الصياغة الموجبة لجميع فقرات مقياس التوهُّم الذى استُخدم بدرجة أساسية فى قياس هذه الشخصية، جعل الأفراد يستجيبون بنعم أو موافق بغض النظر عن محتوى الفقرات. وللتحقق من حدوث هذا التحيز أُعدت فقرات موجبة ولكن ذات معنى معاكس للمعنى الأصلى، فإذا كان محتوى الفقرات محدِّدا رئيسا لدرجات الأفراد سوف يكون هناك ارتباط سالب بين الدرجات الأصلية والدرجات عندما عُكس المعنى الأصلى. غير أنه وجد أن هناك ارتباطًا موجبًا مما يؤكد تأثر الاستجابات بالإذعان أو الموافقة. ويمكن تقليل أثر هذا الاتجاه العقلى فى الاستجابات بالعناية بصياغة الفقرات بحيث لا يشوبها الغموض، وكذلك استخدام عدد متساو من الفقرات الموجبة والسالبة وتوزيعها عشوائيًّا فى ميزان الاتجاه المراد قياسه كأسلوب لتقييم تأثر الاستجابات بهذا الاتجاه العقلى.

(٥) التركيب اللغوى للفقرات: يتعلق هذا التحيز بميزان الاتجاه وصياغة فقراته، ولكنه يؤثر فى استجابات الأفراد، فاستخدام أقسام مثل «موافق» ، «غير موافق على الإطلاق» ، «غالبًا» ، «عادة» ، «نادرًا» يؤدى إلى تفسيرات متباينة من جانب الأفراد، فما يسميه البعض «أحيانًا» ربما يسميه البعض الآخر «غالبًا» وتجمع هذه الأخطاء المنتظمة يؤدى إلى تحيز الدرجات. غير أنه لا توجد أدلة بحثية كافية حول مدى تأثير الفروق الكمية فى معانى مثل هذه المفردات اللغوية فى درجات موازين الاتجاهات.

بعض استخدامات مقاييس الاتجاهات:

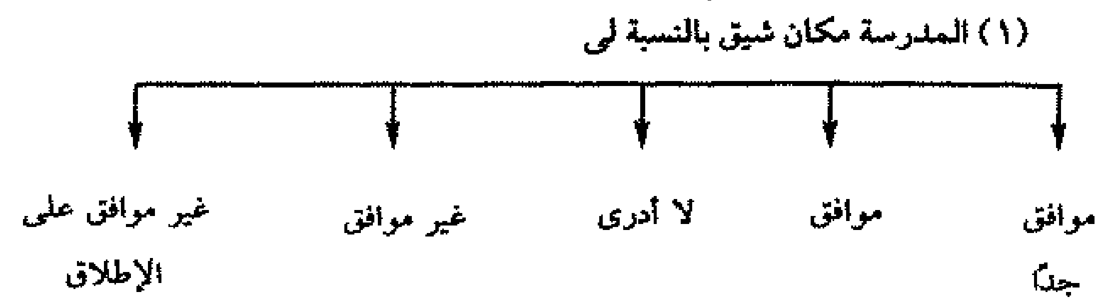
تقويم اتجاهات الطلاب:

ذكرنا فى الفصل السابق أن تقويم الجوانب الوجدانية لدى الطلاب لا يقل أهمية عن تقويم الجوانب المعرفية أو النفسية الحركية، فالأهداف التعليمية للمجالات الدراسية يتعلق بعضها بالجوانب الوجدانية، ويجب أن تؤدى إلى تغيرات فى سلوك الطالب وتثرى تعلمه. فالتعلم والمشاعر الإنسانية مرتبطان ارتباطًا وثيقًا ولا يمكن اعتبارهما ثنائية، فكل طالب يتعلم شيئًا يكون بالطبع لديه اتجاه معين أو شعور ما نحو ما يتعلمه، فإذا كان اتجاه الطالب إيجابيًّا فإنه يثرى ما يتعلمه ويصبح سلوكه بنَّاء، وهذا يتطلب مراعاة العلاقات القائمة بين المعلومات المستقاة من المجالات الدراسية التى تتجمع لدى الطلاب وشخصيته كفرد متميز. فمثلا يجب على المعلم أن يجذب الطالب إلى المجال الدراسى ويجعله واعيًا بأهميته فى البيئة التى يعيش فيها لكى ينمِّى لديه اتجاهًا إيجابيًّا نحو هذا المجال.

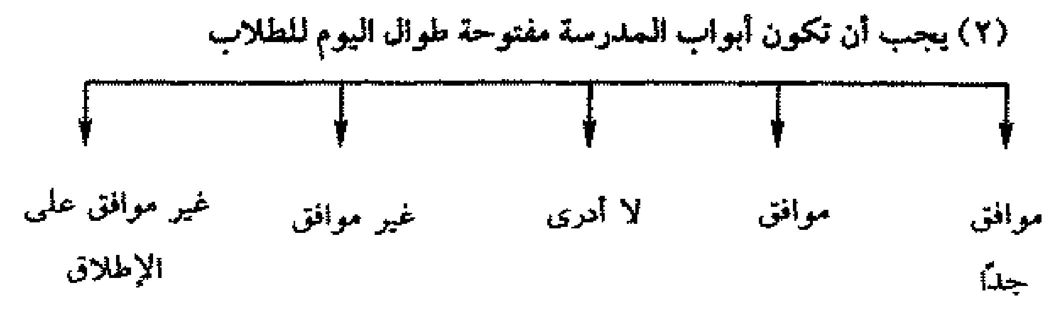
غير أن المشكلة التى تواجه كثيراً من المربين هى كيفية تحديد الأهداف المتعلقة بالاتجاهات وصعوبة قياس هذه الاتجاهات.

فالطريقة المباشرة لإجراء ذلك هى ملاحظة سلوك الطالب فى مواقف محددة، كما يمكن استخدام الموازين المتدرجة واختبارات المواقف، ففى بناء الموازين المتدرجة يعطى الطالب عبارات جدلية أو خلافية معينة يلى كل منها عدداً من الاستجابات التى يمكن أن توضح درجة أو شدة موافقته على كل منها، ومثال ذلك:

(١) المدرسة مكان شيق بالنسبة لى



(٢) يجب أن تكون أبواب المدرسة مفتوحة طوال اليوم للطلاب



ويُطلب من الطالب وضع علامة أو دائرة حول الاختيار الذى ينطبق عليه أو يعبِّر عن رأيه.

ويمكن وضع أوزان رقمية لكل اختيار، وجمع أوزان جميع الفقرات لكل طالب كما فى حالة ميزان ليكرت، وذلك للاستدلال على مدى إيجابية أو سلبية اتجاه الطالب نحو القضية المطلوبة. ويمكن الاقتصار على «موافق» أو «غير موافق» إذا كان ميزان الاتجاه مصممًا لتلاميذ المرحلة الأولى. وهنا يُعطى المعلم الوزن (+ ١) إذا كان التلميذ موافقًا، (ـ ١) إذا كان غير موافق، والعكس فى حالة العبارات المنفية، ثم يجمع أيضًا الأوزان ليستدل منها على مدى إيجابية أو سلبية اتجاه كل تلميذ.

كذلك يمكن قياس الاتجاهات بدرجة أفضل عن طريق اختبارات المواقف، وفيها يُقدم للطالب عبارة تتعلق بموقف من المواقف الواقعية المرتبطة بموضوع أو

قضية موضع الاهتمام يليها عدد من الإجابات أو البدائل المقترحة، وعلى الطالب أن يختار العبارة التي تنطبق عليه، أي تعبر عن رأيه تجاه القضية، ومثال ذلك:

رأيت زميلك يُحرِّض على تعاطي المخدرات، ماذا تفعل؟

(أ) أحاول أن أبيِّن له عواقب ذلك

(ب) أطلب منه أن يكُفّ عن هذا التحريض

(جـ) أحاول معرفة سبب تحريضه على ذلك

(د) أبلغ الأمر لمسئولى المدرسة

(هـ) لا أعطى للأمر أى أهمية

ويفيد تقويم اتجاهات الطلاب فى معاونتهم على تبصُّر ذواتهم مما ييسر نضجهم الشخصى والانفعالى، كما يفيد المعلم فى تنظيم خبرات تعليمية تتعلق بالاتجاهات وتضمين المناهج الدراسية أنشطة لتقديم تغذية مرتدة تساعدهم فى تعرُّف مشاعرهم المتعلقة بموضوعات أو أحداث معينة. وتعرُّف نتائج ذلك للصف الدراسى بأكمله يفيد المعلم فى اتخاذ قرارات تربوية تتعلق بأساليب تنمية اتجاهات إيجابية لدى الطلاب بمعاونة المرشد النفسى والتربوى بالمدرسة من أجل مزيد من الدفء العاطفى فى بيئة التعلم، وتحمُّل مشاعر الآخرين، والمشاركة الجماعية فى اتخاذ القرارات المتعلقة بالأنشطة المدرسية، مما يسهم فى إثراء تعلم الطلاب وإقبالهم على المدرسة، وبذلك يمكن تحقيق الأهداف التربوية المرجوة.

تقويم البرامج التربوية:

تُستخدم مقاييس الاتجاهات في كثير من الأحيان فى تقويم فاعلية البرامج، والمقررات الدراسية والتجديدات التربوية فى المناهج والتقنيات التعليمية. إذ ربما يقوم معلم الدراسات الاجتماعية بتقويم فاعلية البرنامج التعليمى المتعلق بهذا المجال فى تغيير اتجاهات الطلاب نحو كثير من الموضوعات المتعلقة بقضايا ومشكلات مجتمعهم والمجتمعات الأخرى، وبالمثل فى بقية المجالات الدراسية.

كما يمكن استخدام مقاييس الاتجاهات فى دراسة معنويات الطلاب عند التحاقهم بالمدرسة، وأثناء الدراسة وفى نهاية العام. وكذلك يمكن قياس اتجاهاتهم نحو البرنامج التعليمى، وأساليب التعليم وطرائق التدريس، وملاءمة الكتب المدرسية، والممارسات داخل المدرسة، وقضايا ومشكلات مجتمعاتهم المحلية. ويفيد ذلك فى

تعرُّف جوانب القوة وأوجه القصور فى البرامج التربوية والعمل على تطوير هذه البرامج.

التشخيص العلاجي:

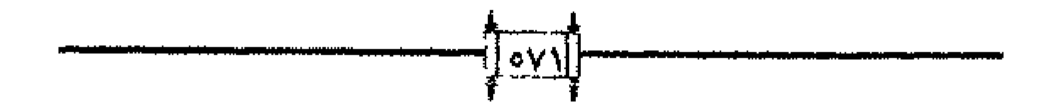
يمكن الإفادة من مقاييس الاتجاهات في تعرُّف أسباب نفور بعض الطلاب من متابعة الدراسة مما قد يؤدى إلى ضعف تحصيلهم أو رسوبهم المتكرر، فالطالب الذى يعتقد أن المدرسة مضيعة للوقت، أو أنها مكان لحشو الأذهان بمعلومات غير نافعة، أو أن المعلمين يستغلون سلطتهم فى جعل الطلاب يمتثلون لآرائهم وتوجهاتهم، من غير المحتمل أن يبذل الجهد والوقت اللازم لتحقيق النجاح والتميز فى دراسته، وهذا يتطلب تحديدًا لاتجاهات مثل هؤلاء الطلاب، وتصميم برامج وأنشطة تعليمية ثرية ومتنوعة تسهم فى تعديل اتجاهاتهم نحو المدرسة والتعلم.

الإرشاد التربوى والانتقاء:

تُستخدم أحيانًا مقاييس الاتجاهات في الإرشاد التربوى للطلاب، فمثلاً بعض موازين الاتجاهات المقننة المتوافرة تهدف لمسح اتجاهات الطلاب نحو معلميهم، ونحو طرائق التدريس والأهداف التربوية ومتطلباتها، ونحو الدراسة بعامة وتوجهاتها، لتحديد السلوك المميز لكل منهم، وتعريفه باتجاهاته ومشاعره نحو مختلف مكونات العملية التعليمية، والتمييز بين الطلاب الذين لديهم اتجاهات إيجابية أو سلبية نحوها.

غير أن موازين الاتجاهات ربما تكون محدودة الفائدة فى مجال انتقاء الأفراد إذا لم يكن موضوع الاتجاه مرتبطًا أو متعلقًا بالأداء فى مهنة أو عمل معين، فالفرد الذى لديه اتجاه اجتماعي أو سياسى معين ربما لا يؤثر فى أدائه لعمل معين، وبالتالى لا يصبح هذا الاتجاه مؤشرًا يُستند إليه فى انتقاء الفرد للالتحاق بهذا العمل. كما أن بعض الأفراد يلجئون كما سبق أن ذكرنا إلى تزييف استجاباتهم فى موازين الاتجاهات وفقًا لمتطلبات الانتقاء، فإذا تقدم فرد للالتحاق بمهنة التدريس مثلا، ربما لا يُظهر عدم رضاه عن العمل مع الأطفال، أو سلبية اتجاهه نحوهم، أو نحو المهنة ذاتها نظرًا لحاجته الماسة إلى الالتحاق بأى عمل، وبالتالى تصبح فائدة استخدام مقياس الاتجاه فى عملية الانتقاء موضع تساؤل.

وعمومًا، فإنه ينبْغى عدم الاقتصار على نوع واحد من مقاييس الاتجاهات، وإنما يمكن استخدام مقاييس متعددة إذا تطلب الأمر ذلك. إذ يمكن استخدام موازين الاتجاهات إضافة إلى مقاييس أخرى غير مباشرة، وفى جميع الأحوال ينبغى مراعاة أوجه قصور كل من هذه المقاييس وارتباطه بالمجال المراد قياسه وبمتطلبات عمليات التقويم والتشخيص والإرشاد والانتقاء.



خلاصة

تُعدُّ الاتجاهات من الجوانب الوجدانية الأساسية، حيث ترتبط بمشاعر الفرد ومعتقداته وحاجاته ودوافعه ورغباته، ويُستدل عليها من بعض الأنماط السلوكية التى تنعكس فى سلوك الفرد، أو من استجاباته اللفظية أو غير اللفظية لمقاييس الاتجاهات بأنواعها المختلفة. وهناك العديد من المفاهيم المتعلقة بمفهوم الاتجاه وتختلف عنه بدرجات متفاوتة مثل الميول والآراء والمعتقدات والقيم والدوافع والعادات والروح المعنوية.

ولا تقتصر الاتجاهات على المكونة الوجدانية وإنما تتضمن تفاعل هذه المكونة مع المكونتين المعرفية والسلوكية بما يجعل الفرد أكثر تهيؤا واستعدادًا لكى يسلك بطريقة معينة نحو موضوع الاتجاه. وتتميز الاتجاهات بخصائص أساسية من أهمها التوجُّه، والشدة، والمركزية، والمرونة، والشمول، والاتساق. وتتعدد طرق وأساليب بناء مقاييس الاتجاهات، فبعضها يتعلق ببناء موازين مقننة تُعد بمثابة أساليب تقرير ذاتى، والبعض الآخر يعتمد على أساليب الملاحظة، والمقابلة الشخصية، أو يعتمد على أساليب غير مباشرة أو تنكُّرية، مثل الأساليب الإسقاطية، والأساليب الفسيولوجية. ومن أهم أساليب بناء موازين الاتجاهات أسلوب الفترات المتساوية ظاهريًا (ثيرستون)، وأسلوب التقدير الجمعى (ليكرت)، والأسلوب التراكمى (جتمان)، وأسلوب تمايز معانى المفاهيم (أوزجود) ، وأسلوب التصنيف الترتيبى Q Sort (ستيفنسون) ، وأسلوب البسْط (كومبس) ، وأساليب قياس العلاقات الاجتماعية، والموازين متعددة الأبعاد، بأنواعها المختلفة. وتختلف هذه الأساليب فى مستويات قياسها وخصائصها المختلفة. وتُستخدم موازين الاتجاهات استخدامات متعددة فى الأغراض البحثية والتطبيقية أكثر من غيرها من المقاييس غير المباشرة، غير أن هناك بعض المشكلات المتعلقة بالاتجاهات العقلية للاستجابات Response Sets ينبغى محاولة التغلب عليها سواء فى تصميم موازين الاتجاهات وصياغة فقراتها، أو فى تقنين عمليات تطبيق هذه الموازين لكى تكون نتائجها صادقة ومتسقة. ولعل تطوير مفهوم الاتجاه وأساليب قياسه فى المجال التربوى يتطلب مزيدًا من تأكيد علاقة الاتجاهات بالتعلم المدرسى واعتباره مدخلا من مدخلات العملية التعليمية، وناتجًا أساسيًا من

نواتجهـا، وما يترتب على ذلـك من قرارات تربوية. وتوجـيه مزيد من البحـوث نحو تطوير أساليب قياس الاتجاهات لبناء موازين تتميز بنوعية جيدة وتتوافر فيها الخصائص السيكومترية، ومحاولة التغلب على المشكـلات الفنية المتعلقة بهذه الموازين التى تقلل من صدقـها وثبات درجاتـها، واستخدام هـذه الموازين فى عمليات التقويـم التربوى والنفسى.

الفصل الثالث عشر

قياس الشخصية

* مقدمة
* قياس الشخصية من منظور تاريخى
* مفهوم الشخصية وبعض نظرياتها
* تصنيف مقاييس الشخصية
* طرق وأساليب قياس الشخصية

(أولا) مقاييس محددة البنية وواضحة الهدف (استبيانات الشخصية)

* استراتيجيات بناء استبيانات الشخصية
* أمثلة لبعض استبيانات الشخصية
* استبيانات تستند إلى أساس نظرى منطقى
* استبيانات تستند إلى التحليل العاملى
* استبيانات تستند إلى أساس إمبيريقى
* استبيانات جبرية الاختيار
* تقييم استبيانات الشخصية

(ثانيا) مقاييس غير محددة البنية وغامضة الهدف (الأساليب الإسقاطية)

* تصنيف الأساليب الإسقاطية
* أمثلة لبعض الاختبارات الإسقاطية
* تقييم الأساليب الإسقاطية

(ثالثا) مقاييس غير محددة البنية وواضحة الهدف (الملاحظة، والمقابلة، ومقاييس مفهوم الذات)

* أساليب ملاحظة السلوك

* الملاحظة فى مواقف واقعية، ومواقف مضبوطة
* مصادر الخطأ فى أساليب الملاحظة
* بعض الاقتراحات للتغلب على أخطاء الملاحظة
* موازين تقدير السلوك
* مصادر الخطأ فى موازين تقدير السلوك، وكيفية التغلب عليها
* أساليب المقابلة
* أنواع أساليب المقابلة
* ميزات وعيوب أساليب المقابلة
* تقييم أساليب الملاحظة والمقابلة

(رابعًا) مقاييس محددة البنية وغامضة الهدف (مقاييس الأنشطة الفسيولوجية، ومقاييس الإدراك)

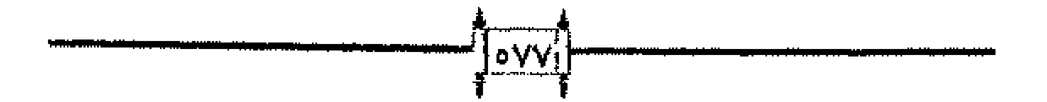
* مقاييس الأنشطة الفسيولوجية
* مقاييس الإدراك
* تقييم الأساليب الموضوعية
* تعقيب عام على طرق وأساليب قياس الشخصية
* استخدامات مقاييس الشخصية
* خلاصة

مقدمة :

عرضنا فى الفصول السابقة أساليب وأدوات قياس التحصيل الدراسى والكفايات والذكاء والاستعدادات الخاصة، أى الجوانب المعرفية، وتتعلق جميعها بقياس الأداء الأقصى للفرد، كما عرضنا أساليب وأدوات قياس الميول والاتجاهات، أى الجوانب الوجدانية، وتتعلق بقياس الأداء المميز للفرد. ونظراً لأن شخصية الإنسان Personality بمفهومها العام تجمع هذه الجوانب وغيرها من المتغيرات المتعلقة بالتفكير، والدافعية، والسمات المزاجية، والمشاعر، والسلوكيات، فى منظومة متكاملة تصف الأداء المميز للفرد أو نمط سلوكه المميز المتسق نسبياً، لذلك فإن أساليب تقييم الشخصية تتناول مجموعة متسعة من الخصائص المعرفية والوجدانية وغيرها من الخصائص التى يُطلق عليها جميعاً متغيرات الشخصية Personality Variables. وهذه المتغيرات تُمكن الفرد من فهم سلوكه وسلوك غيره من الأفراد والتنبؤ به. فحياة الفرد تعد سلسلة متصلة من التفاعلات بينه وبين غيره من الأفراد، ويُصدر من خلال هذه التفاعلات العديد من الأحكام عن شخصيات من يتفاعل معهم فى محاولة التعامل المثمر مع كل منهم. فالجار يحاول فهم شخصية جاره، والوالد يحاول تقييم سلوك وشخصية أبنائه، والمعلم يحاول معرفة الفروق الفردية بين شخصية طلابه، والطبيب يحاول التعرف على شخصية مرضاه حتى يتسنى له معاونتهم فى العلاج، وهكذا.

فمحاولات الأفراد فهم شخصية وسلوك الآخرين تُعد مشاركة من جانب كل منهم فى عملية معقدة تتعلق بتقييم الشخصية الإنسانية. ولكننا بلا شك لا نستطيع أن نعتبر هذه المحاولات أسلوبًا للقياس. ولقد حاول علماء النفس التمييز العلمى الدقيق بين شخصيات الأفراد باتباعهم الطرق والأساليب الكمية فى القياس. فعالم النفس يود معرفة مقدار سمة شخصية معينة لدى فرد ما، أو السمات التى تُميّزه عن غيره من الأفراد، وأحيانًا يود مقارنة فرد ما بفرد آخر، أو مقارنته بمتوسط مجموعة معينة، أو مقارنة سمات فرد الآن بسماته فى وقت سابق. فتحديد مقدار أو إجراء مقارنة يتطلب عمليات قياس.

وبينما يبدأ كثير من الأفراد بإصدار أحكام معينة على غيرهم ممن يتعاملون معهم ثم يُعدّلون من أحكامهم كلما زاد اندماجهم معهم؛ نجد أن عالم النفس يحاول عادة أن تكون أحكامه على الأفراد مبنية على أسس دقيقة باتباعه أساليب علمية فى القياس.



لذلك سوف نتناول فى هذا الفصل مفهوم الشخصية وبعض نظرياتها، وتصنيف المقاييس المتعلقة بهذه النظريات، والأساليب المتنوعة لقياس الشخصية، واستراتيجيات وطرق بناء الأنواع المختلفة من مقاييس الشخصية، وأمثلة متعددة لهذه المقاييس، وما يتعلق بكل نوع منها من مصادر خطأ وكيفية التغلب عليها، ونُقدّم تقييمًا لكل أسلوب من أساليب القياس، وتقييمًا عامًا لهذه الأساليب، ونوضح الاستخدامات المتعددة لمقاييس الشخصية.

قياس الشخصية من منظور تاريخى:

على الرغم من أن الدراسات العلمية للشخصية بعامة لم تبدأ قبل منتصف العقد الثالث من القرن العشرين، إلا أن محاولات تصنيف الأشخاص فى أنماط أو أقسام استنادًا إلى الخصائص الجسمية أو الفسيولوجية كانت سائدة قبل ذلك بزمن بعيد. ونظرًا لأن هذه المحاولات المبكرة تبيّن عدم صحتها، فقد أصبحت غير مقبولة فى وقتنا الحاضر. وقد أعد جالتون Galton أول استبيان للشخصية عام ١٨٨٠ لدراسة التصور، واستخدم هول Hall استبيانًا للشخصية فى دراسته لنمو المراهق، وكذلك استخدم كربلين Kraeplin، وسومر Sommer اختبارات التداعى الحر، وأعد يونج Jong اختبارًا للتداعى الحر عام ١٩٠٥ لقياس الانفعالات غير السوية.

ومما شجع على تطوير مقاييس مقننة للشخصية، الحاجة إلى فحص الخصائص الشخصية للأفراد الملتحقين بالجيش فى الحرب العالمية الأولى. ولعل قائمة وودورث لبيانات الشخصية Woodworth Personal Data Sheet تُعد أول مقياس للشخصية محدد البنية تم بناؤه أثناء هذه الحرب، ونُشر فى صيغته النهائية عقب انتهائها. وهذا يدل على أن نشأة مقاييس الشخصية ارتبطت بضرورات عملية نتيجة تحديات معينة، كما هو الحال فى نشأة مقاييس الذكاء والجوانب المعرفية الأخرى. وتُعد هذه القائمة بمثابة مقابلة شخصية مقننة يجيب فيها الفرد إما بـ «نعم» أو بـ «لا» عن ١١٦ فقرة أو عبارة تتعلق بالسلوك، وكان الهدف منها تحديد الأفراد الذين ربما يصعب عليهم التكيف للمواقف الضاغطة فى الحياة العسكرية. وقد تبين فائدة هذه القائمة كأداة فحص فى ذلك الوقت، غير أنها لم تكن أداة تشخيصية، كما لم تستند إلى أساس نظرى لبنية السلوك سواء فى بنائها أو فى تفسير درجاتها؛ لذلك اتجه علماء النفس الكلينيكى إلى بناء مقاييس أخرى تستند إلى أساليب أكثر تطورًا. ويُعد استبيان برنرويتر للشخصية Bernreuter Personality Inventory من المقاييس التى نالت قبولا واسعًا خلال الأعوام التى تلت الحرب العالمية الأولى، حيث اشتمل على أربع خصائص

وصفية للشخصية هي: النزعات العُصابية، والاكتفاء الذاتي، والانطواء، والسيطرة. وقد اتسع نطاق استخدام هذا الاستبيان لمدة طويلة، ولكن قلَّ استخدامه في وقتنا الحاضر، حيث تلاه العديد من استبيانات الشخصية نتيجة النقد الذي وُجِّه للاستبيانات المبكرة للشخصية.

وبينما كان كثير من علماء النفس في الولايات المتحدة الأمريكية في العشرينيات يقومون بإعداد استبيانات محددة البنية لقياس الشخصية وتتطلب ورقة وقلمًا، كان هناك اتجاه آخر يهدف لبناء مقاييس غير محددة البنية. فقد اهتم هرمان رورشاخ H. Rorschach في سويسرا بوضع أسس المقاييس الإسقاطية Projective Techniques التي تستند إلى فكرة أن الفروق في شخصية الأفراد تتعلق بالفروق في إدراكهم للعالم المحيط بهم.

لذلك بدأ التجريب على مثيرات غير محددة البنية، مثل التصميمات المجردة التي يمكن الحصول عليها من بقع الحبر لكي يستجيب لها الفرد بما يدركه. وقد نُشر اختبار بقع الحبر The Rorschach Ink Blot Test عام ١٩٢١، ويُعد من أكثر الاختبارات المعروفة في قياس الشخصية حتى وقتنا الحاضر، ويُستخدم في التشخيص الكلينيكي. غير أن الاهتمام بهذا النوع من القياس ازداد ببطء، فقد مرت عدة أعوام قبل نقل اختبار رورشاخ إلى الولايات المتحدة الأمريكية بواسطة ليفاى Levy ، وأجريت حوله أول دراسة للدكتوراه عام ١٩٣٢ بواسطة بيك Beck أحد تلاميذ ليفاى، حيث اهتم بالدراسة العلمية لخصائص هذا الاختبار. وقد استثارت هذه الدراسة علماء النفس هناك على الرغم مما قوبلت به الأساليب الإسقاطية من ارتياب وشك ونقد من جانبهم. وعلى الرغم من ذلك قام موراى Murray ومورجان Morgan بإعداد اختبار تفهم الموضوع The Thematic Apperception Test (TAT) عام ١٩٣٥، وهو يختلف عن اختبار رورشاخ في أن بنية مثيراته أكثر تحديدًا من بنية بقع الحبر، إذ يشتمل على صور غامضة تمثل مناظر ومواقف متنوعة، ويُطلب من المفحوص أن يكوِّن قصة تتعلق بهذه المثيرات الغامضة.

وقد نال هذان الاختباران اهتمامًا متزايدًا في أواخر الثلاثينيات وأوائل الأربعينيات، ولعل هذا كان مرجعه إلى وقفة التأمل من جانب علماء النفس فيما يتعلق بالاستبيانات محددة البنية من أجل التغلب على المشكلات المتعلقة ببنائها وتفسير درجاتها. غير أن المقاييس الإسقاطية للشخصية، وبخاصة اختبار رورشاخ لم تحظ بعد ذلك بقبول عام.

ولعل إسهامات مورفى Murphy وجنسن Jensen عام ١٩٣٧ بمؤلفهما بعنوان مداخل للشخصية Approaches to Personality الذى تناولا فيه بالتفصيل مدارس علم النفس، وطب الأمراض النفسية، وتوجيه الأطفال، وكذلك إسهامات ألبورت Allport المتعلقة بمفهوم الذات قد أثرت تأثيراً واضحًا فى التنظير السيكولوجى فى مجال الشخصية.

وقد حدث أثناء الحرب العالمية الثانية تطورات ملحوظة فى أساليب بناء استبيانات الشخصية، فقد أعدّ عام ١٩٤٣ استبيان الشخصية متعدد الأوجه (MMPI)، واستند إلى الأساليب الإمبيريقية فى بنائه. ويُعد هذا الاستبيان من أكثر مقاييس الشخصية استخدامًا فى وقتنا الحاضر ونال قدراً كبيراً من الدراسات والبحوث. وتلى ذلك بناء استبيان كاليفورنيا السيكولوجى (CPI) عام ١٩٥٧ استناداً إلى الأساليب الإمبيريقية أيضًا.

وقد استرشد علماء النفس بالتطورات السيكومترية، وبخاصة أساليب التحليل العاملى Factor Analysis ، وأساليب الاختيار الجبرى أو القهرى Forced-Choice فى بناء استبيانات الشخصية. ويُعد كاتل Cattell، وثيرستون Thurstone، وأيزنك Eysenck، وجيلفورد Guilford من الرواد الذين استخدموا أساليب التحليل العاملى فى مجال قياس الشخصية. وتُعد المحاولات الجادة التى قام بها جيلفورد فى أوائل عقد الأربعينيات، وكاتل فى نهاية العقد من الجهود البارزة فى هذا المجال. فقد أعد كاتل Cattell استبيان العوامل الستة عشر للشخصية (16 PF)، حيث اعتنى عناية كبيرة ببنائه استناداً إلى أسلوب التحليل العاملى.

واهتمت بحوث قياس الشخصية بعد ذلك بتطوير نظرية فى الشخصية أكثر من اهتمامها بأساليب القياس من أجل صياغة تكوينات فرضية استناداً إلى هذه النظرية. لذلك نلاحظ أن مقاييس الشخصية تُستخدم حتى وقتنا الحاضر بكثير من الحرص والحيطة من جانب اختصاصيى علم النفس والتربية نظراً لخصوصية هذا النوع من المقاييس كما سيتضح فى هذا الفصل.

مفهوم الشخصية وبعض نظرياتها :

إن قياس وتقييم الشخصية يُعد مجالا رئيسًا من مجالات القياس التربوى والنفسى، ولكى يمكننا فهم مقاييس الشخصية فهمًا واضحًا ينبغى أن نعرف المقصود بمفهوم الشخصية، ولماذا نحاول قياسها. فعندما نكون بصدد مفهوم أساسى فى مجال السلوك الإنسانى ، فإننا نواجه مشكلة تعدد تعريفات هذا المفهوم. ولقد سبق أن

لاحظنا ذلك فيما يتعلق بمفهوم الذكاء. وينطبق ذلك أيضًا على مفهوم الشخصية، فتعريف الشخصية يختلف باختلاف وجهة نظر علماء النفس ومدارسهم الفكرية والسيكلوجية. وقد استنتج ألبورت Allport من دراسته المسحية المبكرة التى أجراها عام ١٩٣٧ أن هناك ما يقرب من ٥٠ تعريفًا على الأقل لمفهوم الشخصية. وربما يرجع ذلك إلى صعوبة تحديد ما يشتمل عليه هذا المفهوم المتسع الذى يعد من أكثر المفاهيم السيكلوجية غموضًا (Smith & Adams, 1972).

إذ يرى جيلفورد (Guilford, 1959) أن الشخصية تنطوى على جميع خصائص الفرد المستقرة نسبيا، وتتضمن المظهر الجسمانى، والذكاء والاستعدادات، والميول، والاتجاهات، والقيم، والأمراض النفسية، ويرى غيره أن مفهوم الشخصية يقتصر على السمات السلوكية غير المعرفية التى تكشف عن مفهوم الذات، والخُلُق، والنسق الفردى، وبخاصة المكونات الانفعالية والدافعية لسلوك الفرد.

وبذلك يُستبعد الذكاء والاستعدادات والمظهر الجسمانى، غير أن التوجُّهات الحديثة اقترحت ضرورة أخذ الجوانب المعرفية بعين الاعتبار من أجل التوصل إلى نظرية متكاملة للشخصية (Mischel, 1973). وتستند هذه التوجهات إلى أن جميع مظاهر السلوك الذى يكشف عن الشخصية منشأها العقل، ولكى نستطيع متابعة دراسة الشخصية يجب عدم إغفال العقل، كما أن المواقف الموضوعية المحيطة بسلوك الفرد ربما تختلف اختلافًا جوهريًا عن المعنى الذاتى الذى يضفيه الفرد على هذه المواقف، حيث إن هذه المواقف ليست منعزلة عن الفرد.

لذلك لا يوجد تعريف صحيح للشخصية وتعريف خطأ، كما لا يوجد تعريف وحيد صواب، وإنما يعتمد التعريف على استخداماته المناسبة المتعلقة به، فمثلا يذكر ألبورت (Allport, 1961) أن الشخصية هى التنظيم الديناميكى داخل الفرد للأنظمة النفسية الجسمانية التى تحدد سلوكه المميز وتفكيره. ويذكر سندبرج (Sundberg, 1977) أن الشخصية هى مجموعة العمليات التى يستخدمها الفرد أو الأفراد فى تنمية انطباعات وتصورات، واتخاذ قرارات والتحقق من فرضيات تتعلق بأنماط خصائص فرد آخر التى تحدد سلوكه فى تفاعله مع البيئة. ويرى كراون (Crowne, 1969) أن الشخصية تُعد نظامًا من الإمكانات للسلوك، ويُقصد بذلك أننا نلاحظ التنظيم والوحدة فى السلوك، حيث إن سلوك الفرد ليس جزيئيًا أو منفصلاً، وإنما يكون متسقًا عبر الزمن وباختلاف المواقف، فالشخصية تتعلق بالفرد ككل وليس فقط بذكائه أو اتجاهاته أو بعض دوافعه. فهى تتعلق بهذه جميعًا وبغيرها من المتغيرات وبأثر تفاعلاتها فى

السلوك، لذلك يشير كراون إلى أهمية التمييز بين التحليل الكلينيكي لشخصية فرد معين من أجل فهم الشخصية الوظيفية للفرد ككل، وبين المدخل الذى يتبعه الباحثون فى دراسة الشخصية حيث يعتمد على عزل بعض الأحداث السلوكية ودراسة أثر كل منها وكذلك أثرها المشترك فى التجارب متعددة المتغيرات. كما أن هذا التعريف يؤكد الإمكانات للسلوك Potentialities، وليس السلوك فى ذاته، فالاستدلال من السلوك عن مُحركاته وتوجُّهاته أى إمكاناته هو الذى يُمكننا من الفهم العام المتسع للسلوك وبالتالى للشخصية.

وعمومًا فإن علماء النفس يمكن أن يتفقوا على عدد من الخصائص العامة لتعريف مقبول متسع للشخصية، على الرغم من أن استنادهم إلى تعريفات توجِّه دراساتهم يمكن أن تعدِّل أو تتغاضى عن إحدى هذه الخصائص أو بعضها. ولعل التعريف التالى يُقدم إطارًا يفيد فى توجيه الحوار حول مفهوم الشخصية. فالشخصية تشير إلى تكامل جميع خصائص الفرد المستقرة نسبيا فى تنظيم فريد يحدد محاولاته للتكيف مع بيئته دائمة التغير، ويمكن تعديله فى ضوء هذه المحاولات. ويتضمن هذا التعريف العناصر التالية:

(١) تعدد السمات والإمكانات والقدرات التى تشتمل عليها الشخصية.

(٢) انتظام هذه السمات والإمكانات والقدرات فى نسق متكامل.

(٣) تفرُّد الشخصية، حيث تُميِّز بين الفرد وغيره من الأفراد.

(٤) تأثير الشخصية فى العلاقات التفاعلية بين الفرد وبيئته وبغيره من الأفراد.

(٥) اعتبار الشخصية من خصائص الفرد المستقرة نسبيا عبر مدة زمنية طويلة.

وقد حاولت نظريات الشخصية وضع تصورات نسقية لهذا المجال الذى يتميز بالتعقد بطرق مختلفة، لعل أكثرها شيوعًا اقتراح أبعاد أو سمات تتباين حولها الشخصية، أو اقتراح تفسيرات توضح نمو الشخصية ودينامياتها، فالمدخل الأول يركز على طبيعة بنية الشخصية، أما المدخل الثانى فيركز على عمليات نمو الشخصية.

ويختلف توجُّه الدراسات الإمبيريقية فى كل من المدخلين، فالنظريات البنائية تستند إلى نتائج اختبارات الشخصية فى تحديد السمات الأساسية للشخصية، فقد توصلت بعض الدراسات إلى أبعاد ثنائية القطب، مثل الانطواء ـ الانبساط ، الهوس ـ الاكتئاب ، الذكورة ـ الأنوثة، وتوصلت دراسات أخرى إلى أبعاد متعددة للشخصية السوية أو العادية، فالتقرير الموضوعى لشخصية الفرد فى إطار هذا المدخل الذى يتعلق

بالسمات Trait Approach يتطلب تقدير موقعه أو مكانته فى أبعاد متعددة. والشكل أو التكوين العام لهذه التقديرات أو المواقع المتعلقة بسمات الفرد يسمى الصفحة النفسية Psychogram أو البروفيل النفسى Profile . وقد سبق أن أوضحنا ذلك فى الفصل السادس.

أما المدخل الثانى الذى يستند إلى نظريات التعلم ونظريات التحليل النفسى؛ فيركز على تحليل عمليات نمو الشخصية، ويؤكد العوامل الوقتية والموقفية تأكيداً كبيراً، وكذلك يؤكد العمليات المهمة فى التكيف الإنسانى، مثل التعلم والدافعية والصراعات، وربط ميكانيزمات التعميم والتمييز ، وتكوين العادات السلوكية بمشكلات الشخصية.

وبالطبع توجد مداخل أخرى تستند إلى نظريات تجهيز المعلومات Information Processing وغيرها من النظريات المعرفية.

وسوف نوضح هذه المداخل المتعددة فى سياق مناقشتنا لطرق وأساليب قياس الشخصية.

تصنيف مقاييس الشخصية :

اتضح لنا مما سبق أنه لا يوجد اتفاق بين علماء النفس على تعريف مفهوم الشخصية، مما أدى إلى تباين وتعدد النظريات التى حاول كل منها إلقاء الضوء على هذا المفهوم من منظور معين. وقد ترتب على ذلك تباين وتعدد طرق وأساليب قياس الشخصية وما يتعلق بها من متغيرات. والحقيقة أن مجال قياس الشخصية يُعد من أكثر مجالات القياس النفسى والتربوى تنوعاً فى طرقه وأساليبه، وربما يرجع ذلك كما سبق أن أوضحنا إلى ما يتضمنه مفهوم الشخصية من متغيرات ومفاهيم أخرى غير متجانسة. وعلى الرغم من تعدد طرق تصنيف مقاييس الشخصية، إلا أن هناك عناصر مشتركة يمكن أن تجمع بينها، فقياس أو تقييم الشخصية يُعد أسلوباً لجمع المعلومات عن فرد ما، وعلى وجه التحديد يتضمن الملاحظة الموضوعية المنظمة للسلوك تحت شروط محددة فى علاقتها بمثيرات معينة، وللحصول على بيانات أو معلومات عن فرد ما يجب محاولة فهم أثر مختلف مكونات الموقف المتعلق بالمثير على سلوكه، ويتحدد هذا الموقف الذى يواجهه الفرد ويستجيب له بخصائصه الفيزيائية، وبالتعليمات التى تُقدم له، وبالمطلب المحدد.

ونظرًا لاختلاف المتغيرات المرتبطة بعملية جمع البيانات الشخصية عن الفرد، فإن هناك تباينًا واضحًا في طرق وأساليب جمع هذه البيانات. فطبيعة كل من: الموقف، والمثير، والتعليمات المعطاة، والاستجابة المطلوبة، وكذلك كيفية تقدير الدرجات وتحليلها وتفسيرها، تُعد خصائص لعملية قياس أو تقييم الشخصية. إذ يُستند إليها في عملية جمع البيانات وتسجيلها وتحليلها وتفسيرها. لذلك ينبغي اتباع أسلوب منظم في الحصول على هذه البيانات في إطار نظرية معينة أو فروض نظرية واضحة ومحددة المعالم، فاختلاف نظريات الشخصية يؤدي إلى طرق مختلفة لملاحظة سلوك الأفراد. أي أن هناك علاقات تفاعلية بين نظريات الشخصية وطرق وأساليب القياس أو التقييم.

ويرى أرندت (Arndet, 1974) أن نظرية الشخصية التي لا تسهم في تحديد وتيسير وسائل جمع البيانات تُعد نظرية غير ذات جدوى من الناحية العملية والتطبيقية، كما أن البيانات غير المرتبطة بنظرية معينة تكون عديمة المعنى وقليلة الفائدة؛ لذلك ينبغي أن يستند تصنيف مقاييس الشخصية إلى إطار نظري أو فكري واضح، وأن يُراعى الفروق بين مختلف نظريات الشخصية، وتوجهات البحوث السيكولوجية في هذا المجال.

وتتباين طرق وأساليب قياس أو تقييم الشخصية في مكونات عملية جمع البيانات التي أشرنا إليها. فالموقف الاختباري يمكن أن يكون شبيهًا بمواقف الحياة الواقعية، أو يكون محاكاة له، أو يكون موقفًا مختبريًا محدد المثيرات، وكذلك يمكن أن يتضمن الموقف مثيرًا واحدًا أو عددًا كبيرًا من المثيرات محددة البنية أو غير محددة البنية (غامضة). وبعض المواقف الاختبارية يتطلب تسجيل استجابة واحدة وبعضها الآخر يتطلب تسجيل عدد من الاستجابات، وهذه الاستجابات إما أن تكون تقريرًا لفظيًا ذاتيًا، أو استجابات فسيولوجية ، أو أداءً معينًا، أما تقدير الدرجات فيمكن أن يكون موضوعيًا ويمكن إجراؤه بواسطة الحاسوب أو يدويًا، كما يمكن أن يكون ذاتيًا استنادًا إلى أحكام الأخصائيين النفسيين. وفي الحالة الأولى يتم تفسير الدرجات من خلال الصفحات النفسية أو البروفيل Profile الذي نحصل عليه باستخدام برنامج الحاسوب، وفي الحالة الثانية يتم تفسير البيانات تفسيرًا إكلينيكيًا، بواسطة هؤلاء الأخصائيين.

لذلك يمكن تصنيف مقاييس الشخصية في بُعدين أساسيين كل منهما ثنائي القطب (Campbell, 1957; Mehrens & Lehman, 1969) من حيث بنية المثيرات، والهدف من الاختبار. فبنية المثيرات يمكن أن تكون محددة Structured أو غير محددة Nonstructured، والهدف من الاختبار إما أن يكون واضحًا Nondisguised أو غامضًا

Disguised ، وبذلك يمكـن أن تنقسم مقـاييس الشخصيـة إلى أربعة أنواع يوضـحها الشكل التخطيطى (١٣ـ١) التالى:

أنواع مقاييس الشخصية

- مقاييس محددة البنية
  - واضحة الهدف
    - استبيانات الشخصية
      - المدخل المنطقى
      - مدخل التحليل العاملى
      - المدخل الإمبيريقى
  - غامضة الهدف
    - مقاييس الأنشطة الفسيولوجية ومقاييس الإدراك
      - المدخل التجريبى (الموضوعى - السلوكى)
- مقاييس غير محددة البنية
  - واضحة الهدف
    - أساليب الملاحظة والتقدير والمقابلة
      - المدخل الكلينيكى (الفينومونولوجى)
  - غامضة الهدف
    - الاختبارات الإسقاطية
      - المدخل الكلينيكى (العوامل اللاشعورية)

شكل (١٣ـ١) يوضح أنواع مقاييس الشخصية
(من تصميم المؤلف)

ويتضح مـن شكل (١٣ـ١) أن المقاييس محددة البنـية يمكن أن تكون واضحة الهدف كما فى استبيانات الشخصية Personality Inventories ، أو غامضة الهدف كما فى اختـبارات الأداء المقـننة، ومقاييـس الأنشطة الفـسيولوجـية. ويعتـمد تحديد بـنية المقـياس على مـا يُسمح به للـفرد من حرية الاسـتجابة. فـفى المقيـاس محدد البـنية

يستجيب الفرد لعدد محدود من البدائل ويُقرر ما يراه منها منطبقًا عليه. أما في المقياس غير محـدد البنية فإن الفرد يكون حـرًا في استجابته كـما يراها ويُعبِّر عنها. والـمقياس واضح الهدف يعني أن الفرد يكون مُدركًا للغرض الفعلي من تطبيقه عليه، أما المقياس غامض الهدف أو التنكّري، فإن تفسير الفاحص لنتائجه يختلف عما يفترضه المفحوص حول الهـدف من المقـياس، فالمـقياس الذي يُـطبق على أنه يقـيس القدرة الإدراكـية ويستخـدم الفاحص نتائجـه في الاستدلال على الاضطراب النفسي مثلا يكون مقـياسًا محدد البنية، ولكنه غامض الهدف.

وقد سـبق أن أوضحنا ذلـك في الفصل السـابق عند مـناقشتـنا لأساليب قـياس الاتجاهات. ويستند بناء استبيـانات الشخصية إلـى المدخل السيكومتـرى أو المدخل الإمبيـريقي الذي ناقشناه في الـفصل الحادي عشـر فيما يتعـلق بأساليب بناء مـقاييس الميول. أما اختبـارات الأداء ومقاييس الأنشطة الفسيولوجية، فإنـها تستند إلى المدخل الموضوعي السلوكي الذي يـعتمد على الأساليب التجريبيـة في جمع بيانات عن سلوك الفرد في موقف يتم ضبط متغيراته.

وكذلك المقاييس غير محددة البنية يمكن أن تكون واضحة الهدف، مثل المقابلة الشخصـية حيث تهـتم بالجانب الكلـينيكي، وتؤكـد المـدخل الفـينومـونـولـوجي Phenomonological Approach، ويكون الفرد مدركًا لذاته وللبيئة المحيطة به، كما يمكن أن تكون غامضة الهدف مثل الاختبارات الإسقاطية Projective Tests التي تهتم بالجانب الكلينيكي أيضًا ولكن تؤكد العوامل اللاشعورية. لذلك تعتمد هذه الاختبارات على مثيرات غامضة مـثل التصميـمات المبهـمة ، والصور ، والكلمـات، والجمل، والأفعال.

وعلى الرغم من أنه يمكن تصنيف مقاييس الشخصية في أقسام أخرى إضافة إلى هذه الأقسام الأربـعة، أو في أقسام تستنـد إلى أساس مختلف، إلا أن التـصنيف الذي يوضحه شكل (١٣ـ١) يُعـد مناسبًا وكافيًا لتـوضيح المقاييس التي سنتنـاولها في الجزء التالي.

طرق وأساليب قياس الشخصية:

نظرًا لتعدد طرق وأساليب قياس الشخصية التي أمكن تصنيفها في الأقسام الأربعة الموضحة بشـكل (١٣ـ١)، فسوف نتناول كل قسم منهـا على حدة، ونوضح الأساس النظري الذي يستند إليه بناء المقاييس المتعلقة به، ونقدم بعض الأمثلة للمقاييس شائعة الاستخدام في كل قـسم منها، وتقييما لهـذه الأقسام. ونظرًا لأن استبيانات الـشخصية

والاختبارات الإسقاطية تُعد أكثر المقاييس شيوعًا واستخدامًا، فسوف نبدأ مناقشتنا بالقسمين الأول والرابع، ثم ننتقل إلى القسمين الثاني والثالث.

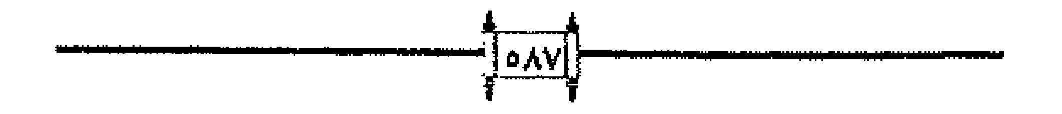
أولاً : مقاييس محددة البنية وواضحة الهدف (استبيانات الشخصية) :

تُعد استبيانات الشخصية Personality Inventories من المقاييس محددة البنية وواضحة الهدف بالنسبة للفرد المستجيب، كما تُعد من المصادر الأساسية للحصول على بيانات ومعلومات تتعلق بالعديد من سمات الشخصية، وتستند هذه الاستبيانات إلى نظريات السمات Trait Theories. فقد أوضح ألبورت Allport أن شخصية الفرد تتكون من تنظيم ديناميكي من السمات التي تُحدد أسلوبه المميز في التكيف مع بيئته. وقد اتفق ألبورت Allport وكاتل Cattell في ضرورة قياس السمات الأساسية التي تُميز بين الأفراد، والتي يعتقدان أنها عوامل محددة داخل الفرد.

غير أن ألبورت يرى أن المدخل المتعلق بصياغة القوانين Nomothetic Approach ليس كافيًا بذاته، ويجب أن يُعزز أو يُستكمل بدراسة ووصف التنظيم الفريد المميز لكل فرد Idiographic Approach. بينما يرى كاتل أن هذا التنظيم أو هذه البنية Structure يمكن قياسها بذاتها باستخدام الأساليب الإحصائية المتقدمة. وفي حين اهتم كاتل Cattell، وجيلفورد Guilford في الولايات المتحدة الأمريكية بمدخل السمات Trait Approach، نجد أيزنك Eysenck في انجلترا اهتم بمدخل الأنماط Types Approach، فمدخل السمات يتضمن تحديد عدد من السمات أو خصائص السلوك المستقرة نسبيا التي تُميز بين الأفراد ولكنها تُمثل نزعة مشتركة بينهم، فالكسل مثلاً خاصة مشتركة بين الأفراد على الرغم من تباين درجاتها لكل منهم.

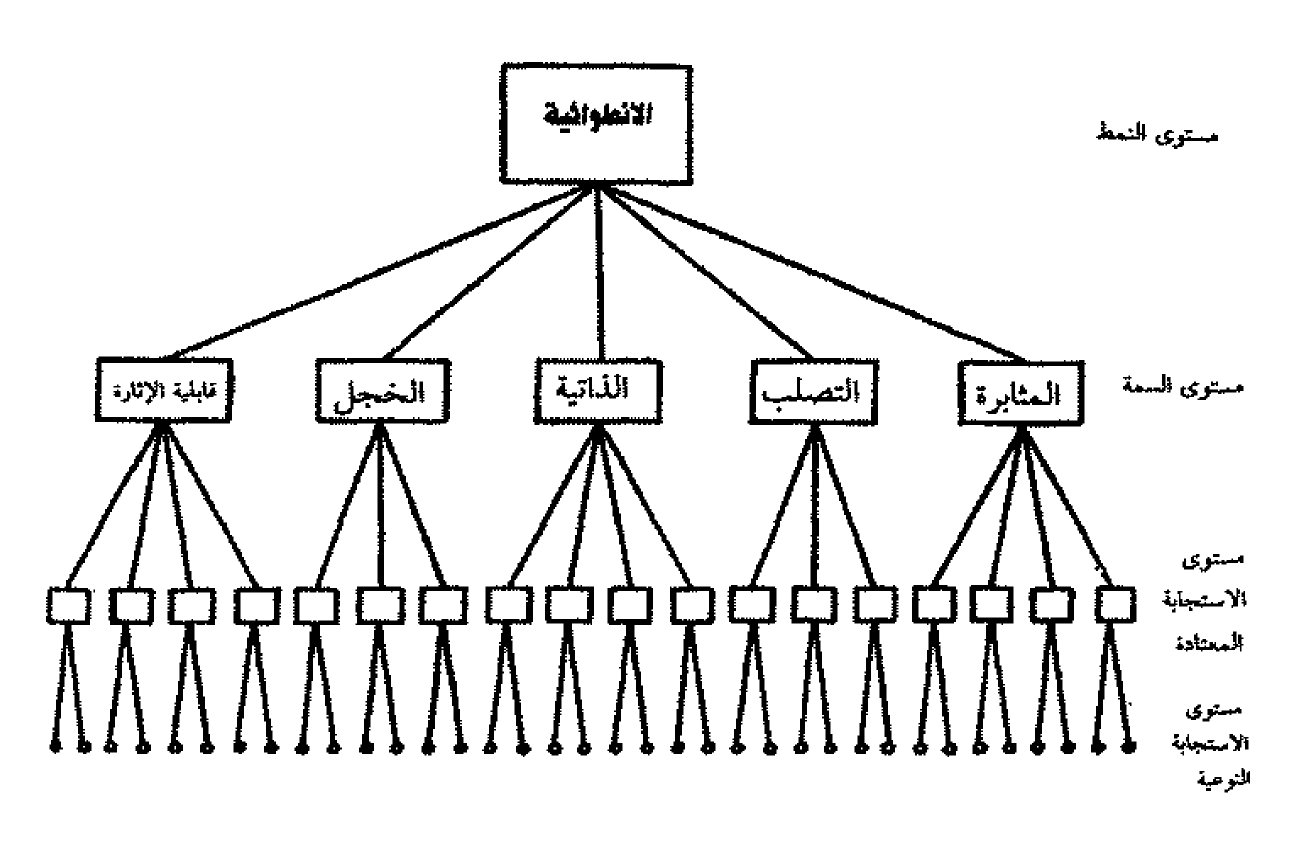
لذلك تهدف نظريات السمات تحديد أهم سمات الشخصية الإنسانية وأكثرها دلالة ، وابتكار أساليب لقياس هذه السمات بدقة لدى كل فرد، أما مدخل الأنماط فيهدف لتصنيف الأفراد في أنماط وفقًا لتجمّعات معينة من السمات، أي أن الأنماط تُعد وحدات وصفية ذات رتب أعلى من السمات.

وقد أجرى أيزنك Eysenck دراسة رائدة عام ١٩٤٧ طبق فيها عددًا كبيرًا من المقاييس على عينة تشتمل على ١٠٠٠٠ فرد من الجنود العاديين والعصابيين، إضافة إلى ملاحظات كلينيكية وجسمانية، وتوصل باستخدام أسلوب التحليل العاملي إلى



عاملين رئيسيين كل منهما ثنائي القطب هما: الانطوائية ـ الانبساطية ، والعصابية ـ الاستقرار.

وقد أجرى دراسة أخرى عام ١٩٥٢ اشتملت على عينات من الأفراد العاديين وأفراد من مرضى المستشفيات العقلية، وتوصل إلى عامل ثالث أطلق عليه عامل الذُهان Psychoticism ويتعلق بجماعة خاصة. وتمثل هذه العوامل الثلاثة المستوى الأعلى في تصنيف أيزنك حيث تدل على أنماط الشخصية. وقد قام بتقسيم كل منها إلى سمات تندرج تحتها استجابات معتادة، وهذه تنقسم إلى استجابات نوعية. أي أن مدخل الأنماط أدى إلى تصنيف هرمي أو بنيوي لسمات الشخصية ، فمثلاً نمط الشخصية المتعلق بالانطوائية Introversion ، وهو المستوى الأعلى، وما يندرج تحته من مستويات أوضحه أيزنك بالشكل التخطيطي (١٣ـ٢) التالي (Eysenck, 1953):



شكل (١٣ـ٢) يوضح النموذج الهرمي لبنية الشخصية فيما يتعلق ببُعد الانطوائية

أما كاتل فقد ميَّز بين السمات المصدرية Source Traits والسمات السطحية Surface Traits، فالسمات المصدرية تتكون من مجموعات من المتغيرات يبدو أنها ترتبط فيما بينها ارتباطًا متسقًا إلى حد ما، بينما تمثل السمات السطحية المتغيرات التى تحدد الأحداث على المستوى السطحي للشخصية ، لذلك تُعد السمات المصدرية أكثر أهمية من السمات السطحية في تفسير السلوك والتنبؤ به؛ لأنها أكثر استقرارًا واتساقًا من السمات السطحية. فالسمات السطحية يمكن ملاحظتها بسهولة في سلوك الفرد، ولكن السمات المصدرية يمكن الكشف عنها فقط باستخدام الأساليب الإحصائية. ويمكن تحديد أو تعرُّف السمات المصدرية باستخدام أسلوب التحليل العاملي للعلاقات القائمة بين السلوك السطحي، كما ميز كاتل بين السمات من حيث كونها عامة أو مُتفرِّدة، موروثة أو متأثرة بالبيئة، وديناميه أو تتعلق بالقدرات، أو المزاج. فالسمات العامة يشترك فيها جميع الأفراد، بينما السمات المتفرِّدة تميز فردًا بعينه، والسمات الديناميه توجِّه الفرد نحو هدف معين، والسمات المتعلقة بقدراته تحدد قدرته على تحقيق الهدف ، وسماته المزاجية تتعلق بالجوانب الانفعالية لنشاطه الموجه نحو الهدف.

وقد استند كاتل على ثلاثة مصادر رئيسة مختلفة في الحصول على البيانات في دراساته هي:

(١) سجل حياة الفرد Life Record (L Data) : وتتعلق بملاحظات الأفراد عن شخصية فرد يعرفونه جيدًا.

(٢) استبيانات التقرير الذاتي Self-Rating Questionnaire (Q Data): وتتعلق بملاحظات الفرد عن نفسه.

(٣) اختبارات موضوعية Objective Tests (T Data) : وتتعلق بأداء الفرد في اختبارات مواقف محددة.

وقد استخدم البيانات المستمدة من هذه المصادر في تكوين مصفوفات معاملات الارتباط بين متغيرات الشخصية لكل نوع من أنواع البيانات الثلاثة، وأجرى تحليلاً عامليًا على هذه المصفوفات، وتوصل إلى ١٦ عاملاً أوليًا أو سمات مصدرية للشخصية، ويتعلق معظم هذه العوامل بسمات مزاجية، فيما عدا عامل يتعلق بالقدرة، وعامل له طبيعة دينامية رأى كاتل أنه يمثل سمة سطحية.

وعلى الرغم من أن كاتل Cattell اعتمد في قياسه لسمات الشخصية على عوامل يتراوح عددها بين ١٦ ، ٢١ عاملا، إلا ان أيزنك Eysenck فضَّل الاعتماد على ثلاثة

عوامل عريضة نسبيًا وغير مترابطة. وقد أشار أيزنك (Eysenck, 1965) في مؤلفه «بنية الشخصية وقياسها Personality Structure and Measurement» أن العـوامل التي توصل إليها كـاتل تحتاج لمزيد من إمعان النظر، وأن نموذجه الهـرمي الذي يشتمل على عـاملين رئيسين يتـميـز على نموذج كاتـل. غيـر أن هناك بعض أوجه القـصور الفنيـة والإحصائية التي تشوب دراسة أيزنك أشـارت إليها بعض الدراسات الأخرى، مما يجعل ملاحظته السابقة تتطلب مزيدًا من البحث. ومع هذا فإنه يبدو أن هناك اتفاقًا فيما بينهما في عاملي الانبـساطية والقلق اللذين يُعدان عـاملين من الرتبة الثانيـة عند كاتل، لذلك نلاحظ أن نظرية كـاتل المتعلقـة بسـمات الشـخصـية تُعـد إطارًا أفاد في بناء عـدد من استبيانات الشخصية.

استراتيجيات بناء استبيانات الشخصية:

يستند بنـاء استبـيانات الشخـصية إلى استـراتيجـيات مماثلة لاسـتراتيجـيات بناء استبيانات الميول التي سبق أن أوضحناها في الفصل الحادي عشر. ومع هذا سوف نلقي مزيدًا من الضوء على هذه الاسـتراتيجيات في مجـال قياس الشخصية. فـقد أشرنا إلى ثلاث استراتيـجيات رئيسة يُستند إلى إحـداها في بناء مجموعات من الفقـرات المتماثلة استرشادًا بأحـكام الخبراء، ويُستند إلى الثانية في بناء مـجموعة من الفقرات التـي تتميز بالاتساق الداخلي، ويُستند إلى الثالثة في انتقاء فقرات على أساس إمبيريقي اعتمادًا على المقارنة بين مجموعتين من الأفراد.

استراتيـجية تستند إلى أحـكام الخبراء: وتعتمد هذه الاستراتـيجية في تجميع الفقـرات ـ بحيث تشكل مـيزان قيـاس ـ على آراء الخبراء، وقـد أطلق عليهـا لانيون، وجودستاين (Lanyon & Goodstein, 1971) وكذلك ويجـينز (Wiggins, 1973)، الاستراتيجية المنطقية النظرية Rational-Theoretical Strategy، وأطلق عليها جولدبرج (Goldberg, 1974) استراتيجية الحدس Intuitive Strategy، بينما أطلق عليها مالوني ووارد (Maloney & Word, 1976) استراتيجية المحتوى Content Strategy.

ويمكن أن تستند أحكام الخـبراء على الحدس أو الإحـساس العام، كـما يمكن أن تستند إلى الاستدلالات المنظمة من نظرية في الشخصية. فالفقرات التي يرون أنها تتعلق بوضوح بما يودون قـياسه يمكن تجميـعها في استـبيان الشخصـية. ولعل بعض المجلات العـامـة تسـتخـدم هذه الاسـتـراتيـجـية فـي تكوين بعض الأسـئلة المتـعلقـة بنواحي

شخصية وتضع لهـا عنوانًا مناسبًا، مـثل «كيف تنجح فى حـياتك العمليـة؟» أو «كيف تكون محبـوبًا من أصدقائك؟» وهكذا. غير أن الاسـتبيانات المبكرة قـد حاولت إسناد الفقرات إلى أساس نظرى، مـثل قائمة وودورث لبيانات الشخصية، واسـتبيان برنرويتر للشخصية اللذين أشرنا إليهما فى مستهل هذا الفصل.

ويمكن إجراء تحليل للفقرات التى اختارها المحكمون للتوصل إلى مجموعة من الفقرات المتعلقة بما يقيسه الاستبيان؛ لذلك فإن الاستبيان الذى يُبنى على هذا الأساس يفتقر إلـى الصدق التجريبى، إذ ربما يتـميز فقط بصدق ظاهرى نظرًا لأنه اعـتمد على الأحكام الذاتية للقائم ببناء الاستبيـان، ومع هذا فإن ذلك يُعد خطوة أولية لإلقاء الضوء على تكوينات فرضية معينة فى مجال الشخصية.

استراتيجية تستـند إلى الاتساق الداخلى للفقرات : وتعتمد هذه الاستراتيجية اعتمادًا أساسيًّا على أسلوب التحليل العاملى Factor Analysis فى انتقاء فقرات تتميز بالاتساق الداخلى. وقد اتضح لنا ذلك فى دراسات أيزنك Eysenck، وكاتل Cattell، وجيلفورد Guilford فى هذا المـجال ، كمـا سبق أن اتضـح لنا ذلك فى بنـاء بعض مقـاييس الذكاء والاستـعدادات الخاصـة والميول. وتهـدف هذه الاستراتيجية لانتـقاء مجموعـة من الفقرات ترتبط فيمـا بينها ارتباطًا مرتفـعًا وترتبط بغيرها من مـجموعات الفقرات ارتـباطًا منخفضًـا. ولعل جيلفورد يُعـد أحد الرواد الذين استخـدموا أسلوب التحليل الـعاملى فى بناء استبـيانات الشخصية، وأعـد استبيان العوامل بالاشتراك مع مارتن Guilford-Martin Inventory (GMI)، كمـا أعد استـبيـان مسح الـسمات المزاجـية بالاشتراك مـع زمرمان Guilford-Zimmerman Temperament Survey (GZTS) . ويقيس هذا الاستبيان عشر سمـات باستخدام ٣٠ فقرة لكل سمة، ووجدوا أن كل مجموعة من هذه الفقرات متسقة داخليًّا.

غيـر أنه يُعاب على أسـلوب التحليل الـعاملى اختـلاف العوامل التـى يمكن أن تُستخلـص من مصفوفات الارتبـاطات بين المتغيـرات بتغير الفقـرات أو عينات الأفراد المستخدمة؛ لذلك نلاحظ أن ثيرستون (Thurstone, 1951) عندما أعاد تحليل بيانات هذا الاستبيان باستخدام عينات أخرى من الأفراد استخلص سبعة عوامل رئيسة بدلاً من عشرة. كما نلاحظ مما سبق أن إيزنك Eysenck استخلص من بياناته عاملين أو بُعدين رئيسين، واستخلص كاتل Cattell ١٦ عاملا. كما أن الاستبـيانات التى تستند إلى هذه الاستراتيجية تتميز فقط بالصدق العاملى Factorial Validity الذى يشير إلى الاتساق

الداخلى للفقـرات، ولكنها تفتقر ـ مثل الاستـراتيجية السابقة ـ إلى الصـدق التجريبى. وكذلك فائدة مثل هذه الاسـتبيانات تكون محدودة فى المجال الكـلينيكى الذى يتطلب استبيانات تُميز بين مجموعات معينة (Goldberg, 1972).

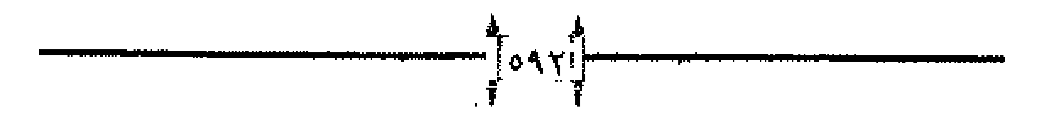
استراتيجية تستند إلى أساس إمبيريقى : يُطلق على هذه الاستراتيـجية أحيانًا استراتـيجية المـجموعات السـمحكة Criterion-Keyed Strategy، أو استـراتـيجـية المجـموعـات المتـناقضة Contrasting Groups Strategy. واستبيانات الشخصية التى تُبنى استنادًا إلى هذه الاستراتيجية يُعتمد فى انتقـاء فقراتها على تمييزها بين مجموعتين أو أكثر يختـلفان فى خاصة أساسيـة معينة، مما يجعـل هذه الاستبيانات تتمـيز بصدق تجريبـى. وقد سبق أن أوضحنا ذلـك بالتفصيل عـند مناقشتنا لاسـتبيانات المـيول فى الفصل الحادى عشر، غير أن الفقرات التى يتم انتقاؤها استنادًا إلى هذه الاستراتيجية لا تكون متجانـسة، بل ربما تتضمن فقرات لا تـتعلق بالسمة المراد قيـاسها مادامت هذه الفقرات تُميز بين مجموعات معينة من الأفراد؛ لذلك تفتقر هذه الاستبيانات إلى أساس نظرى يفسِّر درجاتها.

ومع هذا فإنها تتميز بأن الفرد المستجيب يصعُب عليه تزوير استجاباته للفقرات، ولعل استبيان مينيسوتا المتعدد الأوجه (MMPI)، واستبيان كاليفورنيا للشخصية (CPI) يُعدان من أهم الاستبيانات التى استندت فى بنائها على هذه الاستراتيجية.

ومن هذا يتضح اختلاف الاسـتراتيجيات الثلاث فى الأسس التى تُبنـى استبيانات الشخصية اسـتنادًا إليها، فلكل منها ميـزات وعيوب. ويمكن بناء استبيـانات باستخدام أكثر من استراتيجية للإفادة من ميزات كل منها، إذ يمكن مثلا انتقـاء فقرات تُميز بين أفراد مجمـوعتين متناقضـتين وفى الوقت نفسه تكون مـتسقة داخليًا. أو يمكـن تجميع فقـرات على أساس نظـرى منطقى أولاً وإجـراء تحليل عـاملى على درجـاتها لانتـقاء مجموعات الفقرات المتجانسة ، وهكذا.

أمثلة لبعض استبيانات الشخصية :

نظرًا لكثرة وتنوع استبيانات الشخصيـة المتوافرة ، حيث يبلغ عددها عدة مئات، فسوف نوضح فيما يلى بعض أهم هذه الاستـبيانات شائعة الاستخدام ، وسوف نراعى أن تكون مُستندة فى بنائها على كل من الاستراتيجيات الثلاث التى أوضحناها.



استبيانات تستند إلى أساس نظرى منطقى :

اعتمدت كثير من الاستبيانات المبكرة على قياس سمة واحدة من سمات الشخصية سواء كانت محدودة أو متسعة، لذلك تسمى الاستبيانات أحادية البُعد Uni-dimensional Inventories، وتؤدى إلى درجة كلية واحدة للسمة التى تقيسها. ومن أوائل هذه الاستبيانات قائمة وودورث للبيانات الشخصية عام ١٩١٧ ، واستبيان برنرويتر للاكتفاء الذاتى عام ١٩٣٣. واستند بناء كل منهما إلى أساس نظرى منطقى يتعلق بمحتوى السمة المقاسة. غير أن الاستبيانات الأكثر حداثة اعتمدت على قياس سمات متعددة للشخصية، واستندت فى بنائها إلى التحليل العاملى أو إلى أساس إمبيريقى أو إلى كليهما معًا. وتؤدى إلى عدد من الدرجات الكلية لمجموعة الفقرات التى تشتمل عليها، لذلك تُسمى الاستبيانات متعددة الأبعاد Multidimensional Inventories. وعلى الرغم من اختلاف أساليب بناء نوعى الاستبيانات أحادية البُعد، إلا أن بعض الاستبيانات متعددة الأبعاد اعتمدت على الاستبيانات أحادية البُعد، فمثلا اعتمد استبيان الشخصية متعدد الأبعاد لبرنرويتر على كل من الاستبيانين المشار إليهما أعلاه وعلى استبيانات أخرى أحادية البُعد؛ لذلك سوف نوضح فيما يلى بإيجاز قائمة وودورث ، ثم نُقدم استبيان الشخصية لبرنرويتر.

قائمة وودورث للبيانات الشخصية

Woodworth Personal Data Sheet (WPDS):

تُعد هذه القائمة أول استبيان تقرير ذاتى للشخصية تم بناؤها أثناء الحرب العالمية الأولى، ونُشرت فى صيغتها النهائية عقب انتهاء هذه الحرب مباشرة. واستندت إليها العديد من الاستبيانات التى أُعدّت حديثًا. وكان الهدف منها تحديد الأفراد المجندين الذين يعانون من مشكلات انفعالية تعوق أداءهم لمهامهم العسكرية، وتُعد بديلاً مناسبًا للمقابلة الشخصية الفردية التى كان يجريها الأطباء النفسيون فى ذلك الوقت حيث كان عددهم محدودًا. واعتُبرت بمثابة مقابلة شخصية جماعية، ولكن باستخدام الورقة والقلم، وبذلك مهدت السبيل لبناء فقرات أو عبارات تتعلق بالأعراض المرضية من أدبيات الطب النفسى ومن دراسات الحالات التى تعانى من اضطرابات نفسية، ووُضعت فى قائمة تشتمل على ١١٦ سؤالا يستجيب الفرد لها إما بـ «نعم» أو بـ «لا» ومن أمثلة هذه الفقرات ما يلى:

هل تتلعثم أو يعتقل لسانك أثناء المحادثة؟

هل تعانى من أحلام اليقظة؟

هل منظر بقع الدم يشعرك بالدوار؟

هل تشعر عادة بأن صحتك جيدة؟

هل تشعر بالراحة أثناء النوم؟

هل تشعر بالسعادة معظم الوقت؟

هل تشعر بأن الناس يفهمونك ويتعاطفون معك؟

والدرجة التي يحصل عليها الفرد تشير إلى عدد الأعراض التي يشكو منها ، فالفرد الذى يقرر ذاتيًا كثيرًا من الأعراض يُستبقى لمزيد من الفحص الكلينيكى، أى أن هذه القائمة استُخدمت فى التنبؤ بسوء التكيف، وذلك بتطبيقها على مجموعات كبيرة من الأفراد المجندين.

وعلى الرغم من أن فقرات القائمة اختيرت منطقيًا استنادًا إلى محتواها، إلا أن الفقرات التى استجاب لها ٢٥٪ أو أكثر من عينة الأفراد العاديين استجابة تشير إلى أعراض نفسية عصبية تم استبعادها استنادًا إلى افتراض أنها لا تقيس أعراضًا لسوء التكيف مادام هؤلاء العاديون استجابوا مثل هذه الاستجابات التى ربما لا تكون كذلك إذا أجرى لهم مقابلة شخصية فعلية، وبهذا الأسلوب يقل احتمال الأخطاء الموجبة. وعمومًا فإن الاستجابات لفقرات هذه القائمة وغيرها من استبيانات الشخصية التى تضمنت فقرات مأخوذة منها تتميز فقط بالصدق الظاهرى أو صدق المحتوى الذى تتميز به كثير من اختبارات القدرات والكفايات، على الرغم من صعوبة التحقق من ذلك فى استبيانات الشخصية كما سنوضح فيما بعد.

استبيان الشخصية لبرنرويتر

The Bernreuter Personality Inventory (BPI):

يُعد هذا الاستبيان امتدادًا لاختبار الاكتفاء الذاتى Self - Sufficiency Test الذى سبق أن أعده برنرويتر عام ١٩٣٣ بعنوان : قائمة التفضيل الشخصى Personal Preference Blank التى تضمنت ٦٠ سؤالا يجيب عنها الفرد إما بـ «نعم» أو بـ «لا» أو بـ «؟» وتُقدّم مؤشرات حول مدى اعتماد الفرد على غيره من الأفراد، فالفرد الذى لا يعتمد على غيره يتسم بالاكتفاء الذاتى. واعتمدت هذه القائمة على تجميع الفقرات

وتحريرها، وإعداد محك تقدير الدرجات المتعلق بالاستجابات الثلاث ، وتجريبها على عينة من الأفراد ، وإجراء تحليل الفقرات، ومراجعة القائمة فى ضوء نتائج التحليل ثم إعادة تجريبها ، وإعداد المعايير المئينية. ويُعد اختبار الاكتفاء الذاتى من النوع أحادى البُعد، أما استبيان الشخصية لبرنرويتر فيُعد من المحاولات الأولى لبناء الاستبيانات متعددة الأبعاد. فقد حاول برنرويتر التحقق من صحة افتراض أن سلوك الفرد فى موقف معين ربما يكشف عن سمات شخصية مختلفة ، وعندئذ يمكن تقدير أوزان متباينة لهذه السمات، وبذلك يمكن بناء استبيانات تستخدم فى تحليل سمات متعددة فى آن واحد.

وقد جمع برنرويتر فقرات الاستبيان من عدة استبيانات سابقة، مثل استبيان السيطرة ـ الخضوع الذى أعده ألبورت Allport ، واستبيان الشخصية لثيرستون Thurstone ، واستبيان الانطواء ـ الانبساط الذى أعده ليرد Laird ، واختبار الاكتفاء الذاتى الذى أشرنا إليه، وغيرها من الاستبيانات التى كانت متوافرة. وحاول التحقق من أن مجموعة الفقرات تقيس كل متغير من المتغيرات التى حددها بالدقة نفسها التى تقيس بها الاستبيانات الأصلية هذه المتغيرات.

ويشتمل الاستبيان على ١٢٥ سؤالا يجيب عنها الفرد بـ «نعم» أو بـ «لا» أو بـ «؟» ، وتقيس سمات العُصابية، والاكتفاء الذاتى ، والانطواء ـ الانبساط، والسيطرة ـ الخضوع، والثقة بالنفس، والميل الاجتماعى. ومن أمثلة هذه الفقرات ما يلى:

هل يتحسَّن عملك إذا امتدحك أحد؟

هل تميل إلى ممارسة الرياضة أكثر مما تميل إلى الأنشطة العقلية؟

هل تتجنب عادة استشارة غيرك؟

هل تخجل فى معظم الأحيان؟

هل تشعر أن الناس من حولك يراقبونك؟

وقد بيَّنت معاملات الارتباط بين الدرجات أنها لا تقيس ست سمات مختلفة للشخصية، إذ وُجد أن درجات العصابية ترتبط ارتباطًا مرتفعًا بدرجات الانطواء، ولكن برنرويتر استطاع الحصول على ست درجات وذلك بتعيين أوزان عددية مختلفة للفقرات. وأوضحت نتائج الدراسات التى أجراها أن قيم معامل ثبات درجات كل مقياس فرعى معقولة بالنسبة للاستبيان، أما صدقه فى قياس السمات الست يعتمد على صدق الاستبيانات الأصلية التى جُمعت منها فقرات الاستبيان، ونظرًا لأن صدق بعضها

فى قياس السمة المتعلقة بها كان منخفضًا، فإن صدق استبيان برنرويتر تطلّب المزيد من الدراسات، وقد أُعد مفتاح تصحيح خاصًا بكل مقياس، وكراسة تعليمات توضح كيفية التصحيح، وكذلك صحيفة التقرير السيكولوجى.

ومما هو جدير بالذكر أن هناك بعض الاستبيانات الأخرى متعددة الأبعاد استندت فى بنائها إلى أسس نظرية منطقية، ولعل أكثرها شيوعًا قائمة التفضيل الشخصى التى أعدها إدواردز Edwards Personal Preference Schedule (EPPS) بين عامى ١٩٥٣ ، ١٩٥٩ . ويُستخدم هذا الاستبيان بكثرة فى مراكز الإرشاد النفسى، ونال قدرًا كبيرًا من الدراسات والبحوث، وسوف نتناوله بمزيد من التفصيل فيما بعد.

استبيانات تستند إلى التحليل العاملى :

أشرنا فيما سبق إلى الإطار النظرى لمدخل التحليل العاملى فى بناء الاستبيانات محددة البنية الذى اهتم بـه كل من أيزنك Eysenck ، وجيلفورد Guilford، وكاتل Cattell. فاستبيانات الشخصية المتوافرة فى الوقت الحاضر تفترض أن الأفراد يتميزون بخصائص أو سمات مستقرة نسبيًا ويمكن الكشف عنها فى مواقف متعددة، وتختلف هذه الخصائص من فرد إلى آخر، ويمكن قياسها. ولعل هذا الافتراض يتمثل جيدًا فى الاستبيانات التى يستند بناؤها إلى مدخل أو أسلوب التحليل العاملى Factor Analysis. ومما ساعد فى إجراء العمليات الحسابية والتحليلات العددية التى يتطلبها استخدام هذا الأسلوب توافر الحاسوب الذى يمكن عن طريق برمجياته المتطورة تحليل البيانات متعددة المتغيرات المستمدة من عدد كبير من الاستبيانات وغيرها من مقاييس الشخصية.

وقد اعتمد جيلفورد Guilford فى بنائه لاستبيانات الشخصية على إيجاد العلاقات بين عدد كبير من الاستبيانات المختلفة، وأجرى تحليلاً عامليًا على هذه العلاقات بغرض التوصل إلى الأبعاد أو العوامل التى تشترك فيها هذه الاستبيانات، ومن ثم استخدام هذه العوامل المستخلصة فى بناء استبيان جديد يقيس السمات الأساسية للشخصية . وقد أعد جيلفورد عددًا من الاستبيانات استنادًا إلى هذا المدخل، ثم قام بتحديد العوامل المشتركة بينها واستخدمها فى بناء استبيان مسح السمات المزاجية الذى سنوضحه فيما يلى.

كذلك استخدم كاتل Cattell بمعهد قياس الشخصية والقدرات بولاية إيلينوى الأمريكية هذا المدخل فى بناء استبيان العوامل الستة عشر للشخصية، ولكنه اتخذ توجهًا مختلفًا. فبدلا من إيجاد العلاقات بين استبيانات متعددة استخدم قائمة ألبورت،

وأودبرت (Allport & Odbert, 1936) التى اشتملت على ١٨٠٠٠ صفة تتعلق بالسلوك الإنسانى، ثم حاول اختزالها إلى عدد أقل من السمات التى تُعبر عن معانى المجموعات المختلفة من هذه الصفات، وتوصل إلى ١٧١ سمة (Cattell, 1957). وطلب من عينة من طلاب الجامعات تقييم زملائهم وأصدقائهم فى هذه السمات، وأوجد مصفوفة معاملات الارتباط بينها، ثم أجرى تحليلاً عاملياً على هذه المصفوفة وتوصل إلى ٣٦ عاملاً أطلق عليها كما سبق أن ذكرنا السمات السطحية Surface Traits. وأجرى تحليلاً عاملياً مرة أخرى على هذه العوامل وتوصل إلى ١٦ عاملا أساسياً متميزاً يفسر جميع المتغيرات، أطلق عليها السمات المصدرية Source Traits، واستند إلى هذه العوامل الستة عشر فى بناء استبيانه الذى ستوضحه بعد قليل. ويؤكد كاتل أن هذه العوامل هى الأبعاد الرئيسة التى تُعد ضرورية وكافية لوصف وتفسير الفروق الفردية فى الشخصية التى أشارت إليها الأدبيات السيكولوجية.

استبيان جيلفورد وزيمرمان لمسح السمات المزاجية

: Guilford - Zimmerman Temperament Survey (GZTS)

يشتمل هذا الاستبيان على ٣٠٠ فقرة استمدت من استبيانات سابقة أعدها جيلفورد استناداً إلى أسلوب التحليل العاملى. وتنطوى هذه الفقرات على عشرة عوامل يمثل كل منها ٣٠ فقرة مصاغة صياغة موجبة بدلا من الأسئلة، ويستجيب الفرد لها إما بـ «نعم» أو بـ «لا» أو بـ «؟» وذلك لأن استخدام الأسئلة يزيد من حساسية الفرد فى استجاباته للفقرات، أما الفقرات الموجبة أو الإثباتية تجعل الفرد يستجيب مباشرة. ومن أمثلة هذه الفقرات ما يلى:

أنت شخص نشط فى جميع الأوقات.

يمكنك أن تفكر فى عذر معقول إذا تطلب الأمر ذلك.

نادراً ما تُعيد التفكير فى أخطائك الماضية.

تحب أن تبدأ العمل فى مشروع جديد بحماس شديد.

والسمات التى يقيسها هذا الاستبيان هى:

(١) النشاط العام (الحيوية مقابل الخمول).

(٢) الضبط (الجدية مقابل عدم الاهتمام).

(٣) السيطرة (القيادة مقابل الخضوع).

(٤) المقدرة الاجتماعية (الصداقات المتعددة مقابل قلة الصداقات).

(٥) الاتزان الانفعالى (اعتدال المزاج والتفاؤل مقابل تذبذب المزاج والتشاؤم).

(٦) الموضوعية (الواقعية مقابل الحساسية الزائدة).

(٧) علاقات المودة (اللباقة والاحترام مقابل العداوة والرغبة فى السيطرة).

(٨) الفكر المتأمِّل (الاتزان العقلى مقابل التشتت العقلى).

(٩) العلاقات الشخصية (الثقة بالناس والمؤسسات مقابل الشك والانتقاد).

(١٠) الذكورة (الاهتمام بالأنشطة والمهن المتعلقة بالذكور وكف الانفعالات مقابل الأنشطة والمهن المتعلقة بالإناث والتعبير عن الانفعالات).

ويمكن باستخدام هذه الاستبيانات الحصول على درجات منفصلة لكل من هذه السمات العشر، وكذلك يمكن رسم صفحة نفسية للفرد. وقد اشتمل الاستبيان أيضًا على ثلاثة مفاتيح تكشف عن تزييف الاستجابات أو تهاون الفرد المستجيب، وأُعدت معايير لهذا الاستبيان. وقد وُجد أن قيم معاملات الارتباط بين المقاييس الفرعية العشرة منخفضة إلى حد ما فيما عدا عدد قليل منها، فوسيط هذه القيم بلغ حوالى ٢٢,٠، ولكن قيمة معامل الارتباط بين كل من مقياسى الاتزان الانفعالى والموضوعية بلغت ٦٩,٠ ، وبين كل من مقياسى السيطرة والمقدرة الاجتماعية ٦١,٠ ، وبين كل من مقياسى علاقات المودة والعلاقات الشخصية ٥٠,٠ ، مما يدل على أن السمات التى يقيسها الاستبيان مترابطة إلى حد ما.

وقد تراوحت قيم معامل ثبات درجات المقاييس الفرعية باستخدام طريقة التجزئة النصفية بين ٧٥,٠ ، ٨٧,٠ ، وهى تُعد قيمًا معقولة بالنسبة لمقاييس السمات المزاجية. وقد أُجريت سلسلة من الدراسات للتحقق من الصدق التنبؤى للاستبيان، حيث أوضحت نتائجها أن هناك ارتباطًا منخفضًا موجبًا بين ثلاثة من المقاييس الفرعية والمعدل العام لطلاب الجامعات فى الولايات المتحدة الأمريكية، كما أوضحت أن تقديرات العمل فى المهن التجارية الترويجية ارتبطت ارتباطًا موجبًا دالا بدرجات كل من مقياسى الضبط والفكر المتأمِّل، بينما أوضحت إحدى الدراسات عدم وجود ارتباطات دالة بين أىٍّ من المقاييس العشرة وتقديرات العمل فى المهن الإدارية التنفيذية. غير أن هذا النوع من الصدق كما سبق أن ذكرنا فى الفصل الخامس يعتمد

على المحك الذى يتم اختياره ومدى ملائمته وموضوعيته، وتقديرات العمل تكون فى كثير من الأحيان مشوبة بعنصر الذاتية.

وعمومًا تبيّن أن هذا الاستبيان أكثر صدقًا وفائدة إذا استُخدم مع الأفراد العاديين من الطلاب والعاملين فى مجالات الصناعة مما لو استُخدم مع مرضى العيادات النفسية (Khan, 1962 ; Watley & Martin, 1962). فهذا الاستبيان لم يُصمم بغرض التمييز بين الأسوياء وغير الاسوياء، أو كأداة لتشخيص الأغراض المرضية. ومع هذا توجد بعض المؤشرات على إمكانية الإفادة منه فى هذا الشأن، فقد وُجد أن الأفراد الذُهانيين بعامة Psychotics تكون درجاتهم فى كل من مقياسى المقدرة الاجتماعية، والفكر المتأمِّل منخفضة (Guilford, 1956).

وقد أجرى بنديج (Bendig, 1960) دراسة عاملية أخرى لدرجات المقاييس الفرعية العشر بعد تطبيقها على أربع عينات من مجموعات عمرية مختلفة، واستطاع استخلاص ثلاثة عوامل متعامدة ومتسقة عبر هذه المجموعات وهى: علاقات المودة، والنشاط الاجتماعى، والانبساط ـ الانطواء.

استبيان العوامل الستة عشر لكاتل

: Cattell Sixteen Personality Factor Questionnaire (16 - PF)

يتكون هذا الاستبيان من صيغتين متكافئتين (أ) ، (ب) تشتمل كل منهما على ١٨٧ فقرة، حيث يتراوح عدد الفقرات التى تقيس كلا من هذه العوامل بين ١٠ ، ١٣ فقرة، ويستجيب الفرد لكل فقرة إما بـ «نعم» أو بـ «أحيانًا» أو بـ «لا» ، ويُطلب منه ألا يفكر كثيرًا فى الإجابة، وإنما يُعطى استجابة سريعة كما لو كان فى مواقف حياتية فعلية مماثلة لما تصفها الفقرات، كما يُطلب منه تجنب الاستجابة «أحيانًا» إلا إذا تعذر عليه الاختيار بين البديلين الآخرين، وألا يترك أى فقرة دون إجابة حتى لو لم تنطبق عليه بدرجة جيدة، وأن يتحرى الصدق فى استجاباته قدر الإمكان دون محاولة الاستجابة بما يُرضى الفاحص. ومن أمثلة فقرات هذا الاستبيان ما يلى:

أحب مشاهدة الألعاب الرياضية الجماعية.

المال لا يخلق السعادة.

افضّل الناس الذين يستطيعون تكوين صداقات بسرعة.

والعوامل الستة عشر التى يقيسها الاستبيان هى:

(١) الانعزالية مقابل الانطلاق.

(٢) البلادة مقابل توقُّد الذكاء.

(٣) الانفعالية مقابل النضج الانفعالى.

(٤) الخضوع مقابل السيطرة.

(٥) السكون مقابل التحمس.

(٦) العَرَضية مقابل الشعور بالوجود.

(٧) التهيُّب مقابل الجسارة أو الإقدام.

(٨) الصلابة مقابل الحساسية.

(٩) الثقة بالآخرين مقابل الشك.

(١٠) التمسك بالتقاليد مقابل التحرر.

(١١) البساطة مقابل التعقد.

(١٢) الثقة بالنفس مقابل عدم الأمن.

(١٣) التحفُّظ مقابل التجريب.

(١٤) الاعتمادية مقابل الاكتفاء الذاتى.

(١٥) التراخى مقابل ضبط النفس.

(١٦) الاستقرار الانفعالى مقابل التوتر.

ويحصل الفرد على درجة كلية لكل عامل من هذه العوامل، حيث يُفترض أن هذه الدرجات تعطى صورة عن شخصيته بعامة. ويمكن أن يُجرَى التصحيح يدويًا أو باستخدام برنامج حاسوب مُعد لهذا الغرض، وكذلك رسم صفحة نفسية لدرجات الفرد. ويمكن الرجوع إلى دليل الاستبيان لمعرفة المعنى السيكلوجى لكل من هذه العوامل الستة عشر، حيث تُفسِّر العوامل على أنها سمات الشخصية، وهذا الاستبيان يناسب الأفراد من أعمار ١٦ عامًا أو أكثر. وقد أُعدَّتْ معايير للاستبيان (التساعيات المعيارية Stanines) التى سبق أن أوضحناها فى الفصل السادس، واستندت إلى عينات كبيرة من طلاب المدارس الثانوية والجامعات. ويُعد الاستبيان نوعًا من المقابلة المقننة المنظمة دون أن تكون وجهًا لوجه مع الفاحص، وهدفه محدد وواضح بمعنى

أن الفرد المختبر يعلم أن الاستبيان يقيس الشخصية على الرغم من أن بعض فقراته ربما لا تشير مباشرة إلى ذلك.

وعلى الرغم من الجهود الجادة التى بذلها كاتل ومعاونوه فى بناء هذا الاستبيان، إلا أنه لا يخلو مثل غيره من الاستبيانات محددة البنية وواضحة الهدف من بعض أوجه القصور، ولعل أهمها عدم استقلالية العوامل الستة عشر، وعدم التجانس العاملى لفقرات المقاييس الفرعية، وعدم كفاية المعلومات المتعلقة بالجماعات المرجعية التى استندت إليها المعايير. كما أن قيم معامل ثبات درجات هذه المقاييس باستخدام طريقة التجزئة النصفية كانت منخفضة حيث تراوحت بين ٥٤,٠ ، ٩٣,٠ . وعند ضم الصورتين معًا فى مقياس واحد وتقدير الثبات بطريقة التكافؤ وُجد أن قيمة معامل الثبات حوالى ٥٠,٠ ، وقيمة الثبات الناتجة من إعادة تطبيق الاستبيان بعد حوالى أسبوع كانت أقل من ٨٠,٠ .

غير أن أدلة الصدق الأساسية اعتمدت على أسلوب بناء الاستبيان باستخدام مدخل التحليل العاملى، فكثير من العوامل الستة عشر تناظر العوامل التى تم التوصل إليها من التقديرات والبيانات التجريبية مما يُعزز صدق هذه العوامل. وقد أفادت البيانات المتعلقة بصدق الاستبيان فى اقتراح أفضل طرق استخدام الاستبيان فى مجال الصناعة والمجالات التعليمية والكلينيكية، ولكن البيانات الإحصائية لم تكن كافية بحيث يمكن الإفادة منها فى معرفة ما إذا كانت درجات الاستبيان تتنبأ بالفعل بالسلوك المراد قياسه مما يتطلب مزيداً من الدراسات والبحوث فى هذا الشأن.

ومن الجدير بالذكر أن كاتل أعد استبيانات أخرى للشخصية مماثلة لهذا الاستبيان بحيث تُناسب الأعمار الصغيرة، وهى: استبيان الشخصية للمرحلة الثانوية (١٢ ـ ١٧ عامًا)، واستبيان الشخصية للأطفال (٨ ـ ١٢ عامًا) ، واستبيان الشخصية للمرحلة الدراسية المبكرة (٦ ـ ٨ أعوام).

كما أعد صيغة قصيرة للاستبيان الأصلى ، وصيغتين متكافئتين للراشدين من ذوى الذكاء أقل من المتوسط، وهذا يدل على أن هذه الاستبيانات تناسب تطبيقات عملية عديدة ومتنوعة ، وكذلك تُعد مجالا خصبًا للدراسات والبحوث فى مجال الشخصية. فتوافر مثل هذه الاستبيانات التى تقيس الشخصية بدءاً من عمر ستة أعوام وما فوق يُعد أمراً ضروريًا للدراسات الطولية المتعلقة بنمو الشخصية إذا افترضنا أن هذه الاستبيانات تتميز جميعها بالصدق.

استبيانات تستند إلى أساس إمبيريقي :

أوضحنا فيما سبق أمثلة لاستبيانات تستند في بنائها إلى أساس نظري منطقي يتعلق بمحتوى الفقرات وصدقها الظاهري فيما يقيسه استبيان الشخصية، كما أوضحنا أمثلة لاستبيانات تستند إلى أسلوب التحليل العاملي الذي يُمكننا من انتقاء فقرات متسقة داخليا أي متجانسة. غير أن هناك اتجاهًا آخر يؤكد أهمية بناء استبيانات تتميز بصدق تجريبي وذلك بانتقاء الفقرات التي تميز بين مجموعتين من الأفراد أو أكثر. وقد سبق أن أوضحنا في الفصلين التاسع والحادي عشر أن كلا من مقياسي بينيه للذكاء، واستبيان الميول المهنية لسترونج قد استندا إلى أساس إمبيريقي في انتقاء فقراتهما حيث اعتمدا في ذلك على مجموعات محكة Criterion Groups تم اختيارها بطريقة مستقلة عن أي من المقياسين. وسوف نوضح فيما يلي أمثلة لبعض استبيانات الشخصية المهمة التي استندت أيضًا في بنائها وانتقاء فقراتها إلى هذا المدخل الإمبيريقي Emperical Approach الذي لا يفترض إلا عددًا قليلا من الافتراضات عند انتقاء فقرات الاستبيان، حيث ينبغي فقط أن تكون هذه الفقرات مرتبطة بمحك أو محكات خارجية مناسبة.

استبيان مينيسوتا للشخصية متعدد الأوجه

: Minnesota Multiphasic Personality Inventory (MMPI)

تُعد مشكلة تشخيص سوء التكيف Maladjustment من المشكلات المهمة التي شجّعت علماء النفس على بذل مزيد من الجهد في بناء استبيانات الشخصية. فلقد اتضح لنا أن وودورث Woodworth أعد قائمة البيانات الشخصية لتحديد المشكلات الانفعالية للأفراد المجندين أثناء الحرب العالمية الأولى، وتبعها عدد آخر من الاستبيانات المماثلة. ولعل أهمها جميعًا استبيان الشخصية متعدد الأوجه الذي يُطلق عليه (MMPI) وأحيانًا (Mult) فهو يُعد من استبيانات التقرير الذاتي الأكثر شيوعًا واستخدامًا، والدليل على ذلك أنه أُجري عليه ما يزيد عن ٧٠٠٠ بحث ودراسة (Buros, 1988)، وصدرت حوله كتب مرجعية في عدة مجلدات مثل: An MMPI Handbook (Dahlstrom , Welsh & Dahlstrom, 1960 , 1972 , 1975) ، وتزداد الدراسات والبحوث حوله عامًا بعد عام.

ونال أيضًا نصيبًا وافرًا من النقد والتعليق سواء من الأوساط السيكلوجية أو على صفحات الجرائد والمجلات الأمريكية ، ومع هذا فإنه لا يزال يُمثل اتجاهًا أساسيا في بناء استبيانات الشخصية (Reynolds & Sundberg , 1976) . وقد أعدَّ هذا الاستبيان

هاثاواى Hathaway وهو عالم نفس كلينيكى، وماكنلى McKinley وهو عالم فى الطب النفسى العصبى بجامعة مينيسوتا الأمريكية عام ١٩٤٢ ، حيث تبينت لهما الحاجة إلى بناء استبيان صادق يمكن استخدامه فى التشخيص الكلينيكى لمرضى العيادات النفسية.

ولم يستندا فى بنائه إلى نظريات معينة تتعلق ببنية الشخصية، وإنما استندا إلى المدخل الإمبيريقى ، فقد بدأ العالمان باختيار ما يقرب من ألف فقرة من بطاقات الفحوص الكلينيكية، وكتب الطب النفسى ، وأساليب الفحص العصبى، ومقاييس الاتجاهات الشخصية والاجتماعية التى سبق بناؤها، ثم قاما بتطبيق هذه الفقرات على حوالى ٢٠٠ من المرضى المصابين بأمراض عصبية نفسية فى مستشفيات جامعة مينيسوتا، وكذلك على حوالى ١٠٠٠ من الأفراد العاديين من الممرضات وأقارب المرضى، وعلى عدة مئات من الطلاب المتقدمين للالتحاق بالجامعة، وعلى عينة من سكان إحدى المدن التابعة لهذه الولاية.

وقد تم انتقاء ٥٥٠ فقرة اشتمل عليها الاستبيان فى صيغته النهائية على أساس تمييزها بين هؤلاء المرضى من مثل الفُصاميين والمكتئبين، وبينهم وبين الأفراد العاديين. وقد زاد عدد هذه الفقرات إلى ٥٦٦ فقرة فى الطبعة المعيارية الأكثر شيوعًا واستخدامًا التى صدرت عام ١٩٤٧ (Dahlstrom et al., 1972) ، كما صدرت صيغة مختصرة للاستبيان عام ١٩٥٤ اشتملت على ٤٢٠ فقرة فقط (Olson, 1954) ، وتناولت فقرات الاستبيان مدى متسعًا من الموضوعات المتعلقة بجوانب مختلفة للشخصيه، مثل : النواحى الصحية، والاتجاهات الجنسية والاجتماعية والدينية والسياسية والأسرية والمهنية، والمخاوف المرضية، والحالات الانفعالية المختلفة، والروح المعنوية، وما يتعلق بالذكورة والأنوثة، واتجاه المفحوص نحو الاستبيان.

وُصنفت هذه الفقرات فى عشرة مقاييس فرعية كلينيكية، وأربعة مقاييس فرعية تُستخدم فى التحقق من مدى صدق استجابات المفحوص.

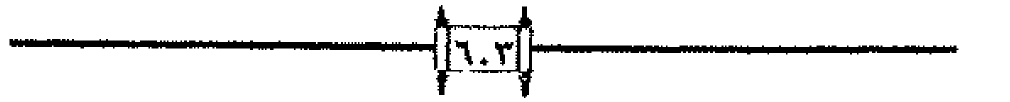
والمقاييس الكلينيكية هى:

(١) توهُّم المرض.

(٢) الاكتئاب.

(٣) الهستيريا التحولية.

(٤) الانحراف السيكوباتى.



(٥) الذكورة ـ الأنوثة.

(٦) البارانويا.

(٧) السيكاثينيا (الوهَن النفسى).

(٨) الفُصام.

(٩) الهوس الخفيف.

(١٠) الانطواء الاجتماعى.

أما مقاييس التحقق من صدق الاستجابات فهى:

(١) المقياس (؟) : وتُقَدَّر الدرجة بعدد الفقرات التى لا يستطيع المفحوص الإجابة عنها بـ «نعم» أو بـ «لا»، وينبغى أن تكون هذه الدرجة أقل ما يمكن.

(٢) مقياس الكذب (ل): يتضمن الفقرات المتعلقة بأمور مقبولة اجتماعيا إلا أنها لا تنطبق على الناس فى عالم الواقع، مثل «لا أقول الصدق دائمًا» . فعلى الرغم من أن الإجابة المعتادة تكون بـ «نعم» إلا أن الإجابة المقبولة اجتماعيًا تكون بـ «لا» . وتُستمد درجة هذا المقياس من مثل هذه الفقرات، فالدرجات التى تزيد عن درجة معينة تشير إلى تعمُّد المفحوص تزييف استجاباته.

(٣) مقياس الشيوع أو عدم الشيوع (ف) : يتضمن الفقرات التى يجيب عنها ١٠٪ أو أقل من الأفراد بالصورة التى تُصحَّح بها، وتعتمد الدرجة على الفقرات التى يُجيب عنها المفحوص بحيث يُظهر نفسه بمظهر غير سوى لأسباب معينة يراها لصالحه أو ربما لعدم فهمه تعليمات الاستبيان ، أو ربما لأن لديه اضطرابًا فعليا.

(٤) مقياس التصحيح (ك): يعكس الموقف الدفاعى للمفحوص فيما يتعلق بالكشف عن مشكلاته، ومحاولة إنكارها لإظهار نفسه بمظهر سوىٍّ، ويبدو أن درجات هذا المقياس تُفيد فى زيادة صدق المقاييس الكلينيكية التشخيصية السابقة.

ومن أمثلة فقرات هذا الاستبيان ما يلى:

تُقلقنى الضوضاء بسهولة.

أعمل تحت ضغوط شديدة.

نومى متقطِّع ومضطرب.

لا أقول الصدق دائمًا.

يبدو أنه لا يفهمنى أحد.

أحب والدى.

أشعر عادة بأن الحياة ذات قيمة.

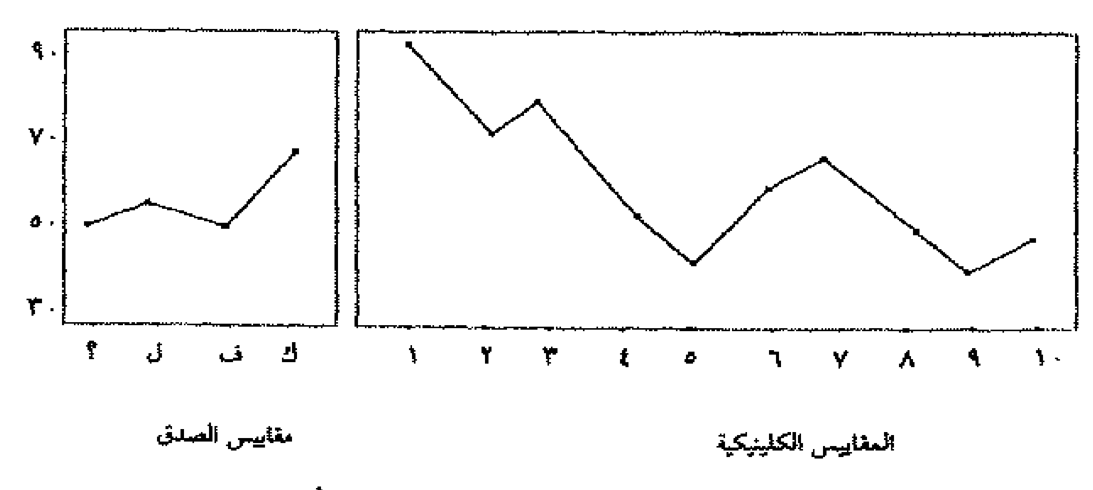
أشعر بالخوف عندما أنظر إلى أسفل من مكان مرتفع.

أشعر بالسعادة معظم الأوقات.

ويتطلب الاستبيان أن يجيب المفحوص إما بـ «صواب» أو بـ «خطأ» أو بـ «لاأدرى» ويدوِّن إجاباته على ورقة إجابة خاصة. وبعد تصحيح الاستبيان وتقدير درجات كل مقياس فرعى تُحوَّل الدرجات الخام إلى درجات تائية T Scores أوضحناها فى الفصل السادس.

ويتم رسم صفحة نفسية (بروفيل نفسى) Profile للمفحوص ، مع ملاحظة أنه أصبح لا يفضَّل استخدام أسماء أو عناوين المقاييس الكلينيكية الفرعية، وإنما تُستخدم الأعداد من (١) إلى (١٠) ، والرموز (؟) ، (ل) ، (ف) ، (ك) لمقاييس الصدق، وذلك لأن هذه المقاييس الكلينيكية أصبح لها معنى مختلف عن المعنى الأصلى الذى كان مقصودًا ، فالأرقام تركز على المعنى الواسع للمقياس الفرعى وليس على عنوانه.

ويوضح شكل (١٣ـ٣) التالى مثالا لصفحة نفسية لفرد عُصابى:

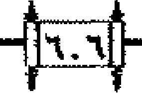


شكل (١٣ـ٣) يوضح صفحة نفسية لفرد عُصابى

ويتضح من شكل (١٣ـ٣) أن الصفحة النفسية اعتمدت على الدرجات التائية التى متوسطها ٥٠، وانحرافها المعيارى ١٠. وكلما كانت إجابات المفحوص غير عادية أو شاذة ارتفعت الدرجات التائية، فالدرجة التائية التى تزيد عن ٧٠ فى أى مقياس على حدة ربما لا تدل على اضطراب متعلق بما يشير إليه مقياس فرعى معين. غير أن تفسير الصفحات النفسية لا يُعد أمرًا بسيطًا ، حيث يعتمد ذلك على تحليل نمط الارتفاعات والانخفاضات فى جميع المقاييس الفرعية Pattern Analysis وليس على الارتفاع المطلق لدرجة كل مقياس على حدة. فالبروفيل أو النمط يتضح من عدد الدرجات المرتفعة فى المقاييس الفرعية. فمثلاً أعلى الدرجات فى شكل (١٣ـ٣) هى درجات المقاييس الفرعية الثلاثة الأولى، لذلك تسمى مقاييس العُصابية Neurotic Triads. ويُقارن هذا النمط بالمفحوصين الآخرين من ذوى نمط ١ ـ ٣ ـ ٢ ، وكذلك بالدلالات الكلينيكية المتماثلة لهؤلاء المفحوصين. وبالطبع ينبغى أولا عند تفسير الصفحات النفسية الاهتمام بدرجات مقاييس الصدق الأربعة لتعرُّف مدخل المفحوص فى استجاباته لفقرات الاستبيان.

وقد اهتمت كثير من الدراسات ببحث الأنماط المختلفة للدرجات، ودلالاتها الكلينيكية، فقد حاولت هذه الدراسات الربط بين الأنماط الشائعة أو المشتركة وبين عبارات وصفية تشخيصية، وتصميم برامج حاسوب تفسِّر الدلالات الكلينيكية لكل من هذه الأنماط آليًا (Graham, 1977)، فالتفسير كما ذكرنا ينبغى أن يعتمد على تحليل عدة مقاييس فرعية فى آن واحد، مما يُعد عملية معقدة تتطلب استخدام الحاسوب، حيث يُجرى هذا التحليل والتفسير بسرعة ودقة وكلفة منخفضة؛ لذلك تقوم بذلك بعض شركات الحاسوب فى الولايات المتحدة الأمريكية، حيث يتم تحليل وتفسير ما يزيد عن عشرة آلاف صفحة نفسية شهريًا فى أحدها للمفحوصين بهذا الاستبيان (Butcher & Dahlstrom, 1980) ، فبرنامج الحاسوب يرسم الصفحة النفسية، ويدون أسفلها عبارات وصفية تفسِّر أنماط الدرجات المتعلقة بالمقاييس الفرعية لكل مفحوص وفقًا لقواعد محددة لتقدير الدرجات وبيانات ترميزية مصممة للبرنامج. وعلى الرغم من أن هذا يُعد توجهًا جديدًا نحو التفسير الموضوعى لدرجات استبيانات الشخصية، إلا أن قبول هذا التوجه يعتمد على الاقتناع بفكرة الموضوعية فى تفسير هذه الاستبيانات، حيث يرى البعض أن هذا التفسير ينبغى أن يستند إلى الأحكام المستبصرة للأخصائيين الكلينيكيين.

ويُستخدم الاستبيان فى التشخيص الكلينيكى لمن هم فى مستوى المرحلة الثانوية وما بعدها، وهو متوافر فى صيغته الأجنبية على شكل بطاقات منفصلة مدون على كل



منها إحدى الفقرات لكى يسهل تطبيقه فرديـا، كما أنه متوافر فى كتيِّب أو على شرائط تسجيل بغرض التطبيق الجماعى، وكذلك توجد صيغة عربية لـهذا الاستبيان، وتُرجم أيضا إلى اللغة الأسبانية.

وقد وُجِّه بعض النقد إلى هذا الاستبيان لانخفاض ثبات درجات بعض المقاييس الفرعية التى اشتمل عليها، حيث بيَّنت بعض الدراسات أن قيم معامل الثبات بطريقة إعادة التطبيق على الأفراد العاديين (معامل الاستقرار) تراوحت بين ٠,٦٠ ، ٠,٩٠ بفاصل زمنى أسبوع واحد، وتقل هذه القيم إذا زادت المدة الزمنية الفاصلة. وقد برَّر مؤلفا الاستبيان انخفاض الثبات بأن الأساليب السيكومترية المتعارف عليها فى تقدير الثبات لا تصلح لاستبيانات الشخصية، وحُجتهما فى ذلك أن كثيرًا من سمات الشخصية متغيرة، وأن البيانات المستمدة من إعادة التطبيق تُعد مؤشرًا لتباين السمات وليس تقديرًا لثبات الاستبيان، حيث يمكن توقُّع هذا التباين فى السمات لدى من يعانون أمراضًا نفسية، وبخاصة من يتلقون علاجًا.

فقد أوضح بيرجن (Bergin, 1971) أن المقياس الفرعى الثانى (مقياس الاكتئاب) بخاصة تتغير درجاته عقب تلقىّ المرضى علاجًا نفسيا فاعلاً، وكذلك وجد أن قيم معامل الثبات بطريقة التجزئة النصفية كانت منخفضة حيث تراوحت بين ٠,٠٥ ، ٠,٩٦ ووسيطها ٠,٧٠ ، وهذا يؤثر على تفسير الصفحة النفسية للمفحوص لتأثرها بالأخطاء العشوائية وعدم استقرار درجاتها؛ لذلك اقترح بعض الباحثين ضرورة انتقاء فقرات جديدة تحل محل بعض فقرات الاستبيان التى تُسهم فى انخفاض الثبات، وذلك فى الدراسات المستقبلية.

كذلك أوضحت الدراسات أن صدق الاستبيان كان موضع تساؤل ، وذلك لقلة عدد أفراد عينة المرضى (أقل من ٥٠ فردًا فى كل مجموعة كلينيكية) التى استُخدمت فى انتقاء فقرات الاستبيان وعدم تمثيلها لهؤلاء المرضى، ولعل ذلك ربما أدى إلى تعيين بعض الفقرات فى المقاييس الفرعية استنادًا إلى عوامل الصدفة (Sax, 1974) . كما أن الصدق التلازمى أى مدى تمييز الاستبيان بين المجموعات الكلينيكية والأفراد العاديين لم يكن مُرضيًا، وأكد ذلك مؤلفا الاستبيان. فمما لا شك فيه أن بناء الاستبيان استنادًا إلى الأساليب الإمبيريقية يعتمد اعتمادًا كبيرًا على طبيعة المجموعات المحكة Criterion Groups ، والتشخيص الكلينيكى لا يكون ثابتًا ويختلف من مستشفى إلى آخر، ومن عيادة نفسية إلى أخرى، وذلك لاختلاف الأساس النظرى الذى يُستند إليه فى التشخيص (Kline , 1976) . غير أن الدراسات والبحوث الكثيرة التى أجريت حول

هذا الاستبيان (ما يزيد عن ٧٠٠٠ دراسة) تُقدِّم دليلا عن صدق التكوين الفرضي للاستبيان، وربما كان هذا أحد أسباب استمرار استخدامه حتى وقتنا الحاضر على الرغم من افتقاره إلى أدلة سيكومترية رصينة تؤيده.

ومن بين أوجه النقد الأخرى تداخل الفقرات في أكثر من مقياس فرعي، أي تكرار كثير من الفقرات في أكثر من مقياس، فمثلا المقياس الفرعي المتعلق بالفُصام (المقياس ٨) يشتمل على ٧٨ فقرة، ١٦ فقرة منها فقط غير متكررة، ولعل هذا قد أدى إلى قيم مرتفعة للارتباط بين المقاييس الكلينيكية الفرعية حيث تراوحت بين ٠,٦٤ ، ٠,٨٧ بحسب عدد أفراد العينات المستخدمة (Dahlstrom & Welsh , 1960). لذلك اهتمت بعض الدراسات بإجراء تحليل عاملي لهذه الارتباطات، وتوصلت إلى عاملين فقط يفسِّران تباين درجات المقاييس العشرة، أحدهما عامل عام ربما يتعلق بالقلق Anxiety أو الفُصامية Neuroticism، ويتأثر تأثراً كبيراً بالاتجاه العقلي للمفحوصين من حيث مراعاة القبول الاجتماعي Social Desirability. ولعل الارتباط المرتفع بين درجات المقاييس الفرعية ، وما توصلت إليه دراسات التحليل العاملي يجعل صدق تحليل أنماط الاستجابات في الصفحات النفسية موضع تساؤل.

كذلك أكدت كثير من الدراسات (Dahlstrom et al. , 1972; Hathaway & McKinley, 1967) أهمية مراعاة الخصائص الديموغرافية للمفحوصين عند تفسير الصفحات النفسية، مثل العمر ، والنوع، والموطن، والتعليم، والمستوى الاجتماعي الاقتصادي وغير ذلك، فما تدل عليه درجات الاستبيان ترتبط بهذه العوامل ارتباطًا واضحًا، حيث تبيِّن أن نمط واحد لأكثر من صفحة نفسية ربما يختلف معناه باختلاف هذه الخصائص.

وكما أوضحنا في الفصل الثاني، فإن المعايير التي أقرتها الجمعية الأمريكية لعلم النفس (APA) أكدت ضرورة إعادة تقنين الاختبارات والمقاييس التربوية والنفسية كل عدة أعوام، ونظراً لأن هذا الاستبيان أُعدَّ منذ زمن بعيد، فإن تغير نوعية الحياة وتعقدها أدى إلى تغير واضح في التعبيرات الانفعالية والسلوك، وهذا يجعل من الصعب تفسير نتائج بعض المقاييس الفرعية التي يشتمل عليها الاستبيان، مثل مقياس الذكورة والأنوثة، والفُصام، مما يتطلب تحديث معايير هذه المقاييس، وكذلك استحداث معايير للمجموعات العمرية المختلفة. لذلك تُنظم الندوات الدولية المتخصصة التي تُعقد كل عام تقريبًا في الولايات المتحدة الأمريكية حول المقاييس النفسية حلقة دراسية تتناول المُراجعة الفنية لهذا الاستبيان (Butcher, 1981) .

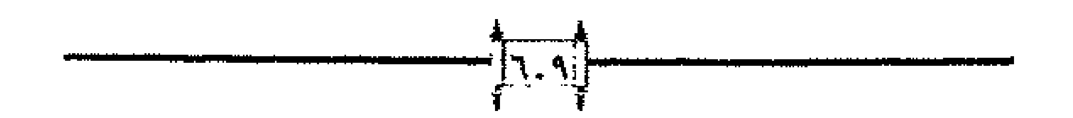
وعلى الرغم من تعدد أوجه النقد التي وجِّهت إلى هذا الاستبيان، إلا أن هذا لا يقلل من قيمة المعلومات التي يقدمها للأخصائي الكلينيكي، والتي يمكن اعتبارها فروضا يمكن أن يتحقق منها استنادًا إلى مصادر بيانات أخرى. فالاستبيان يتميز بشيوع استخدامه وسهولة تطبيقه، وتصحيحه وتفسير نتائجه في ضوء التقنيات المعاصرة، بل يُعد من أكثر استبيانات الشخصية استخدامًا في التشخيص الكلينيكي، والعلاج النفسي، والبحوث السيكلوجية، ويفيد في أغراض الوصف واتخاذ القرارات. كما أثر تأثيرًا واضحًا في تطوير إجراءات التقييم الموضوعي للشخصية استنادًا إلى الأساليب الإمبيريقية بعد ما كان التركيز في الماضي على الأساليب النظرية المنطقية. وتضمن مقاييس فرعية للتحقق من صدق الاستجابات، وتُعد جزءًا متكاملا من إجراءات التقييم والتفسير. كما أن نظم ترميز المقاييس الفرعية ودلالاتها واستخدام الحاسوب في رسم الصفحات النفسية وتفسيرها يُعد من التطورات التي تسهم في موضوعية التصحيح والتفسير.

ولعل ما يميز هذا الاستبيان أن كثيرًا من الاستبيانات الأخرى تم إعدادها باستخدام فقرات أو إجراءات مستمدة منه. ومن أمثلتها استبيان كاليفورنيا السيكلوجي الذي سنوضحه فيما يلي حيث يُطبق على الأفراد العاديين، ومقياس القلق الظاهر لتايلور الذي استُخدم أولا في الأغراض البحثية ثم أصبح يُستخدم في قياس القلق، واستمدت فقراته مباشرة من استبيان الشخصية متعدد الأوجه.

استبيان كاليفورنيا السيكلوجي

California Psychological Inventory (CPI):

يمكن اعتبار هذا الاستبيان صيغة مماثلة لاستبيان الشخصية متعدد الأوجه (MMPI) الذي ناقشناه ، ولكنه يُطبق على الأفراد العاديين وليس على الحالات المرضية كما في الاستبيان الأول. وقد أعدَّ هذا الاستبيان جوخ H. G. Gough، ويشتمل على ٤٨٠ فقرة من نوع الصواب أو الخطأ استُمد نصفها من الاستبيان الأول، أما النصف الثاني فقد أعد خصيصًا لهذا الاستبيان بحيث يناسب الخصائص الشخصية للمراهقين والراشدين. ويشتمل الاستبيان على ١٨ مقياسًا فرعيا تقيس أبعاد الشخصية العادية (السوية) Normal Personality ، وقد صُنفت هذه المقاييس في أربعة أقسام هي: فاعلية العلاقات الاجتماعية التفاعلية، الضبط الذاتي، التوجه الأكاديمي، الاتجاه نحو الحياة (Gough, 1968) ، ويشتمل الاستبيان على ١٨ مقياسًا فرعيا تتعلق بهذه الأقسام الأربعة ، وهذه المقاييس هي:



(أ) مقاييس العلاقات الاجتماعية التفاعلية (النفوذ الاجتماعى، والثقة بالذات):

(١) السيطرة (٤) الحضور الاجتماعى

(٢) المكانة الاجتماعية (٥) تقُبل الذات

(٣) المقدرة الاجتماعية (٦) الإحساس بالكيان

(ب) مقاييس الضبط الذاتى (التنشئة الاجتماعية ، والنضج، وتحمل المسؤلية):

(٧) تحمل المسئولية (١٠) التحمُّل

(٨) التنشأة الاجتماعية (١١) الانطباع الجيد

(٩) ضبط النفس (١٢) الانتماء للجماعة

(جـ) مقاييس التوجُّه الأكاديمى (إمكانات التحصيل، والكفاءة العقلية):

(١٣) الإنجاز عن طريق الامتثال

(١٤) الإنجاز المستقل

(١٥) الكفاءة العقلية

(د) الاتجاه نحو الحياة (الأسلوب العقلى، ونمط الميول):

(١٦) الاتجاه النفسى ـ العقلى

(١٧) المرونة

(١٨) الأنوثة

ومن أمثلة فقرات هذا الاستبيان ما يلى:

أحب الشعْر.

يتوقَّع منى الناس الكثير.

حياتى الأسرية سعيدة دائمًا.

أشعر أحيانًا بأننى على وشك التمزق.

لم أحاول مطلقًا تعمُّد الكذب.

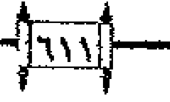
أحيانًا أكون فضوليا.

وتُستخدم مقاييس الإحساس بالكيان، والانطباع الجيد، والانتماء للجماعة فى التحقق من الصدق. وقد أُعد المقياسان الأول والثانى من استجابات الأفراد العاديين الذين يُطلب منهم تزييف هذه الاستجابات لإظهار أنفسهم بمظهر غير سوى أو سوى على الترتيب، بينما درجة المقياس الثالث عبارة عن عدد الاستجابات الشائعة.

وقد تم إعداد ١١ مقياساً من بين المقاييس الباقية عن طريق مقارنة استجابات مجموعات مختلفة من الأفراد، بينما المقاييس الأربعة الأخرى (الحضور الاجتماعى، وتقبل الذات، وضبط النفس، والمرونة) أعدّت استناداً إلى صدق محتواها. أى أن صدق معظم المقاييس استند إلى محكات اجتماعية تتعلق بالحياة الواقعية، وذلك بهدف قياس أبعاد مهمة للشخصية وموائمة لواقع الحياة الاجتماعية. فالمحكات التى استُخدمت هى: المعدّل التحصيلى العام فى المقررات الدراسية لطلاب العينة، وتقديرات الأقران، والمشاركة فى الأنشطة المصاحبة. ولعل هذا يُعد من أهم ميزات هذا الاستبيان، حيث يؤكد إمكانية استخدام الأسلوب الإمبيريقى فى قياس الشخصية العادية (السوية).

وقد استُخدمت أيضاً معايير الدرجات التائية لكل مقياس فرعى، وبيّنت نتائج الدراسات أن هناك ارتباطات مرتفعة نسبيا بين هذه المقاييس، وكذلك أوضحت الدراسات العاملية أن هناك عاملين يفسّران التباين فى درجات كل من مقاييس القسم الأول المتعلقة بالعلاقات الاجتماعية، ومقاييس القسم الثانى المتعلقة بالضبط الذاتى. كما أن ثبات الدرجات لا يختلف كثيراً عن استبيان مينيسوتا (Megagee, 1972) حيث تراوحت قيمه باستخدام طريقة إعادة التطبيق بين ٠,٥٧ ، ٠,٧٧ . ومع هذا فإن الاستبيان يفيد فى كثير من المجالات التى تتطلب تقييم الشخصية، بل إن هذا الاستبيان ربما يعد من أفضل استبيانات الشخصية المتوافرة حاليا، وقد أجريت حوله كثير من الدراسات للتحقق من صدقه، وتمت مراجعته وأُعيد تقنينه عدداً من المرات أكثر من أى استبيان آخر استند إلى أساس إمبيريقى.

فمعاييره اعتمدت على استجابات ٦٠٠٠ من الذكور، ٧٠٠٠ من الإناث من أعمار ومستويات اجتماعية مختلفة، وعلى الرغم من أن صدق كل مقياس فرعى على حدة كان منخفضًا، إلا أنه يمكن استخدام درجات بعض المقاييس الفرعية فى معادلة انحدار للتنبؤ ببعض المحكات، مثل تقديرات الطلاب، والتسرّب ، وجُناح الأحداث وغيرها من المحكات المناسبة.



استبيانات جبرية الإختيار:

تهدف هذه الاستبيانات لضبط أحد الاتجاهات العقلية Response Set التي تؤثر في استجابات الأفراد في معظم استبيانات التقرير الذاتي وتقلل من صدقها، وهي مراعاة القبول الاجتماعي (SD) Social Desirability. وسوف نعرض فيما يلي مثالا لأهم هذه الاستبيانات التي حاول مؤلفها التغلب على هذه المشكلة باستخدام فقرات جبرية الاختيار Forced Choice Format، وقد سبق أن أوضحنا هذا النوع من الفقرات في استبيانات الميول في الفصل الحادي عشر.

قائمة إدواردز للتفضيل الشخصي

:Edwards Personal Preference Schedule (EPPS)

يُعد أسلوب الاختيار الجبري مدخلا مختلفًا في بناء استبيانات الشخصية، ففي معظم هذه الاستبيانات ينزع الأفراد إلى إظهار أنفسهم بمظهر مرغوب فيه، فعندما تكون هناك استجابة واحدة مقبولة اجتماعيا، فإن الأفراد يختارونها دون الاهتمام بمحتوى الفقرة. وقد اهتم إدواردز بمشكلة مراعاة القبول الاجتماعي (SD) التي تؤثر في دقة درجات استبيانات الشخصية، وحاول التغلب عليها في قائمة التفضيل الشخصي التي اعتنى عناية كبيرة ببناء فقراتها، وذلك بأن يكون الفرد مُجبراً على أن يختار إحدى فقرتين متكافئتين في درجة القبول الاجتماعي (Edwards, 1959) . وتُعد هذه القائمة من القوائم المبكرة التي صُممت لتقييم شدة الحاجات Needs التي افترضها موراي (Murray, 1938) بجامعة هارفارد في كتابه « استكشافات في الشخصية Explorations in Personality» حيث وضّح فيه نظامًا للحاجات يشتمل على ١٥ حاجة شخصية Personality Needs مستقلة نسبيًا. وقد قام إدواردز ببناء مجموعات من الفقرات يتعلق محتواها بهذه الحاجات. وتشتمل القائمة على ٢٢٥ زوجًا من الفقرات استنادًا إلى نظرية موراي في الشخصية، وبذلك استطاع إدواردز تجنب المداخل الذاتية التي لا تستند إلى إطار نظري في بناء الفقرات لكي تتميز بصدق المحتوى. وقد حاول تقدير كل فقرة من حيث «مراعاة القبول الاجتماعي» وكوّن منها ثنائيات متوازية في هذا الشأن، ويُطلب من الفرد أن يختار من بين كل زوج الفقرة التي تنطبق عليه، أي تُعبّر عن حاجاته الشخصية بدرجة أفضل، وبذلك لا يستطيع تزييف استجاباته.



فالفرد يكون مُجبرًا على الاختيار بين حاجتين Needs ، أى يختار واحدة على حساب الأخرى فى ضوء محتوى كل منهما ، وبذلك يمكن تمثيل انتقاء الفقرات على أحد المقاييس بالنسبة لانتقائها على مقياس آخر، مما يؤدى إلى درجات تُنسب إلى إطار مرجعى فردى Ipsative Scores التى سبق أن أوضحناها بالتفصيل فى الفصل الحادى عشر. أى أن شدة كل من هذه الحاجات Needs لا يُعبَّر عنها بصورة مطلقة، وإنما فى علاقتها بشدة حاجات الفرد الأخرى. فالتركيز هنا يكون على الفرد ذاته وليس على معايير الجماعة المرجعية؛ لذلك يمكن أن يحصل فردان على درجات متساوية فى الاستبيان، ومع هذا ربما يختلفان اختلافًا واضحًا فى شدة حاجاتهما. وتحويل هذه الدرجات الفردية إلى معايير مئينية ربما لا يكون مناسبًا، بل ومربكًا ويؤدى إلى صعوبات فى تفسير وتحليل الدرجات إحصائيًا.

والحاجات الشخصية التى تتضمنها القائمة هى:

(١) الإنجاز: تخطِّى الصعاب ، بذل أقصى جهد، إنجاز شىء متميز.

(٢) الانقياد: الامتثال بدلاً من القيادة، تقبُّل الآخرين ومدحهم، المسايرة.

(٣) النظام: الحاجة للترتيب، التنظيم، التخطيط المسبق، المدخل المنظم.

(٤) الاستعراض : الحاجة إلى تركيز انتباه الآخرين نحوه، وتحقيق ذلك عن طريق المظهر الخارجى أو استخدام العبارات اللفظية.

(٥) الاستقلال الذاتى: تجنُّب الالتزامات المفروضة، عدم التقيُّد، الاستقلالية، التحرر.

(٦) الانتماء: الولاء، الإخلاص، تقديم العون، الأُلفة.

(٧) التدخل الاجتماعى: تحليل سلوك ودوافع الآخرين، الفهم من خلال تحليل الذات والآخرين.

(٨) تقديم العون: تشجيع الآخرين ومعاونتهم، تعاطف الآخرين فى أوقات الشدة.

(٩) السيطرة: إدارة الحوار، الإشراف والريادة، التأثير، اتخاذ القرارات للآخرين.

(١٠) الاستكانة: الشعور بالنقص، الحاجة إلى العقاب، الحاجة إلى الاعتراف بالخطأ.

(١١) الحاجة إلى التربية الأساسية: معاونة الأصدقاء والمحتاجين، الصفح والتسامح، المودة والكرم والتعاطف.

(١٢) التغيُّر: أداء أنشطة جديدة ومختلفة ، التجريب، القيام برحلات، تجنب الأعمال الروتينية، تكوين صداقات جديدة.

(١٣) التحمُّل: العمل بجد، إنهاء الأعمال المطلوبة، عدم التشتت.

(١٤) الجنسية الغيْرية: المشاركة في جميع أنشطة الجنس الآخر.

(١٥) العدوانية: المخالفة، النقد المتفتح، الانتقام ، التوبيخ.

ومن أمثلة فقرات هذه القائمة ما يلي:

(١) (أ) أحب أن أعاون أصدقائي عندما تعترضهم مشكلة.

(ب) أحب بذل أقصى جهدي في أي عمل يوكل إليّ.

(٢) (أ) أحب أن أوضح ما يتراءى لي عن الأشياء.

(ب) أحب أن أصفح عن أصدقائي الذين يُغضبونني أحيانًا.

(٣) (أ) أشعر بالانقباض إذا فشلت في موقف معين.

(ب) أشعر بالتوتر إذا تحدثت أمام مجموعة من الأفراد.

وإلى جانب هذه المقاييس الفرعية الخمسة عشر قدم إدواردز مقياسًا لاستقرار الصفحة النفسية، ومقياسًا لاتساق الاستبيان يشتمل على ١٥ زوجًا من الفقرات المتطابقة في شكلها وموزعة عشوائيًا في الاستبيان، وبذلك يكون عدد مرات الاختيار المتطابق لكل فرد يمكن تحويله إلى مئينيات استنادًا إلى بيانات معيارية ، كما يمكن تحليل اتساق استجابات الفرد بإيجاد الارتباط بين درجاته في الفقرات الزوجية والفقرات الفردية لكل مقياس فرعي.

وقد استُمدت معايير القائمة من ١٥٠٠ طالب وطالبة بالجامعات ، ٩٠٠٠ من الراشدين من ٤٨ ولاية أمريكية، وأُعدَّت معايير مئينية منفصلة لكل من هاتين المجموعتين، وكذلك لطلاب المدارس الثانوية.

وبيّنت الدراسات أن قيم ثبات درجات المقاييس الفرعية وذلك بإعادة تطبيق الاستبيان بفاصل زمني قدره أسبوع واحد تراوحت بين ٠,٧٤ ، ٠,٨٨ كما هو موضح بدليل القائمة، وقيم الثبات بطريقة التجزئة النصفية تراوحت بين ٠,٦٠ ، ٠,٨٧ . أما الارتباطات بين المقاييس الفرعية فقد تراوحت قيمها بين -٠,٣٤ ، ٠,٤٦ ، وهي قيم أقل من نظيرتها في استبيان كاتل أو استبيان مينيسوتا.

وعلى الرغم من اهتمام إدواردز بضبط متغير مراعاة القبول الاجتماعي في انتقائه لفقرات القائمة، واستناده إلى العلاقة المنطقية بين محتوى كل فقرة وبين تقسيم الحاجات الشخصية الذي افترضه موراي كمحك وحيد لهذا الانتقاء، فإن هذا يؤكد تركيزه على ضبط هذا الاتجاه العقلي على حساب الجوانب الأخرى لصدق القائمة. والدليل على ذلك أن الدراسات المتفرقة التي حاولت التحقق من الصدق يصعب تفسير نتائجها نظرًا لأنها لم تأخذ بعين الاعتبار النظام المرجعي الفردي الذي تستند إليه درجات القائمة، فمتوسط قيم الارتباطات بين المقاييس الفرعية كان سالبًا، ومتوسط ارتباط جميع هذه المقاييس بأي متغير خارجي تقترب قيمته من الصفر (Hicks, 1970) مما يجعل استخدام معامل الارتباط في التحقق من صدق القائمة لا يكون مناسبًا.

ومن بين أوجه النقد الأخرى التي وُجِّهت لهذه القائمة ضرورة التمييز بين مراعاة القبول الاجتماعي ومراعاة القبول الشخصي. كما أنه ربما لا يكون هناك اتفاق بين قيم مراعاة القبول الاجتماعي التي نحصل عليها باستخدام كل فقرة على حدة بالقيم التي نحصل عليها من ثنائيات الفقرات، فمراعاة القبول الاجتماعي للفقرة الواحدة ربما تتأثر بالمضمون الذي تُعبّر عنه. وللتحقق من هذا الافتراض تم الحصول على قيم جديدة لمتغير مراعاة القبول الاجتماعي لثلاثين زوجًا من هذه الفقرات، وتبيّن أنها مختلفة اختلافًا ملحوظًا عن القيم الأصلية التي سبق تحديدها، والتي تم مزاوجة الفقرات استنادًا إليها (Corah et al. , 1958) . كما أنه برزت تساؤلات تتعلق بما إذا كان استخدام فقرات جبرية الاختيار قد أدت بالفعل إلى ضبط متغير مراعاة القبول الاجتماعي (Feldman & Corah, 1960 ; Wiggins 1966) . وقد اقترحت أنستازي (Anastasi, 1976) ضرورة مراجعة هذه القائمة للتغلب على أوجه قصورها الفنية وبخاصة نمط الفقرات وتفسير الدرجات، وكذلك إجراء دراسات صدق تعتمد على تحليل نمط الدرجات التي تستند إلى إطار مرجعي فردي.

ومن الجدير بالذكر أن إدواردز أعد مقياسًا آخر أكثر طموحًا من هذه القائمة أطلق عليه استبيان الشخصية لإدواردز Edwards Personality Inventory (EPI) ، ويشتمل على ٣٠٠ فقرة مدونة في خمسة كتيبات ، واستخدم أسلوب التحليل العاملي

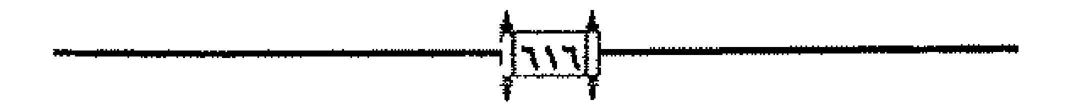
وتقديرات متغير مراعاة القبول الاجتماعي في بناء الاستبيان حيث يقيس ٥٣ بُعدًا من أبعاد الشخصية، بمعدل ١١ إلى ١٥ بُعدا في كل كتيب، وتتطلب الاستجابات أن تكون إما بـ «نعم» أو بـ «؟» أو بـ «لا»، وتمت صياغة الفقرات بحيث تصف الآخرين ويُبدي الفرد رأيه حول ما تصفه الفقرة عن نفسه لزيادة صدق الاستجابات. ويستغرق تطبيق كل كتيب ٤٥ دقيقة تقريبًا، وإجمالي زمن تطبيق الاستبيان حوالي ثلاث ساعات؛ لذلك كانت عينة المعايير متغيرة وغير ممثلة حيث اقتصرت على ٢٠٣ طالب، ٣٢٩ طالبة بجامعة واشنطون وعدد أقل من بعض المدارس الثانوية، ووُجد أن قيمة الثبات باستخدام طريقة كيودر - ريتشاردسون ٠,٨٠ تقريبًا، ولكن لا توجد دراسات كافية تتعلق بصدق الاستبيان.

تقييم استبيانات الشخصية:

بعد عرضنا الموجز للعديد من الأمثلة لاستبيانات الشخصية التي تستند في بنائها إلى أساليب مختلفة وتقيس سمات متعددة ومتباينة، وتُستخدم في أغراض متنوعة، سوف نوضح فيما يلي بعض الملاحظات العامة على هذه الاستبيانات وما يتعلق بها من مشكلات:

(١) تعدد المعاني التي تشير إليها الفقرات: تشتمل استبيانات الشخصية على فقرات محددة البنية يصوغها المؤلف في عبارات معينة، وهنا تبرز المشكلة اللغوية في صياغة العبارات وما تحمله من معنى ربما يختلف من فرد مستجيب إلى آخر، أو يختلف عما يقصده المؤلف. فمثلا إحدى عبارات استبيان الشخصية متعدد الأوجه كالتالي: «والدي كان رجلاً طيبًا» ، وعندما يجيب الفرد عن هذه الفقرة سوف يتبين له أنها تناسب الأفراد الذين تُوفِّي والدهم، أو ربما يُفسرها على أن والده ليس كذلك.

وعبارة مثل «هل عادة تُوجِّه الحوار في المواقف الجماعية؟» فالفرد عليه أن يقرر ما تعنيه المواقف الجماعية ، وهل هي مواقف أسرية أم مواقف في العمل أم مع الأصدقاء؟ وهل تتعلق بمواقف رسمية أم غير رسمية؟ وهل المقصود بكلمة «توجِّه» المبادأة بالحديث، أو التحدث معظم الوقت، أو التعليق على الحوار؟ وهل المقصود بكلمة «عادة» معظم الأحيان، أم في جميع الأحيان. أم دائمًا؟ فصياغة عبارات استبيانات الشخصية تعد أكثر صعوبة من صياغة فقرات مقاييس الجوانب المعرفية أو المهارية، وتحتمل تأويلات وتفسيرات متباينة من جانب المختبرين.



(٢) **انقسام الاستجابات:** يستجيب الفرد لفقرات الاستبيان إما بـ «نعم» ، أو بـ «؟» ، أو بـ «لا» ، وهذا النمط المُقيَّد لأقسام الاستجابة لا يسمح بوصف الإطار المرجعي الذي استند إليه الفرد في استجابته. وأحيانًا يضطر الفرد لاختيار الاستجابة «نعم» على الرغم من أنها لا تصف على وجه التحديد ما يراه الفرد عن نفسه، لذلك يلجأ بعض الأفراد إلى تبرير استجاباتهم في عبارات لفظية يكتبها على هوامش ورقة إجابة استبيان الشخصية.

(٣) **عدم معرفة الفرد لدوافعه وسلوكه:** تفترض استبيانات الشخصية أن الفرد يعرف جيدًا دوافعه وسلوكه الماضي والحاضر بحيث يمكنه أن يستجيب بصدق لفقرات الاستبيان. غير أن هذا ربما يعتمد جزئيا على مدى اضطرابات الشخصية، فكلما كان الفرد يعاني من اضطرابات مرضية شديدة تكون استجاباته المتعلقة بسماته الشخصية غير دقيقة، نظرًا لأنه لا يكون واعيًا بدرجة كافية بدوافعه وسلوكه. كما أن الاستجابات لبعض فقرات الاستبيان ربما تتطلب معلومات لم يلاحظها الفرد من قبل، أو ربما تتطلب منه إصدار أحكام أو تفسيرات تفوق قدراته، وهنا يتطلب التمييز بين التفسيرات أو الاستدلالات وبين الحقائق. ففقرة مثل : « هل تفضِّل العمل الفردي على العمل الجماعي؟» يستطيع أي فرد عادي أن يُجيب عنها بيسر وسهولة، بينما تُوجد فقرات تتطلب من الفرد اتخاذ قرار معين يستند إلى فحص ذاتي لم يتحقق منه سابقًا لتفسير سلوكه، فالفقرات يكون لها مضمون ظاهر ومضمون كامن، وهذا المضمون الأخير هو الذي يعد صادقًا. فالفقرة «هل كرهت والدك أو والدتك في وقت ما؟» يجيب عنها الفرد عادة بـ «لا» مع أن ذلك ربما لا يكون صحيحًا بالنسبة له، ولكنها ربما تكشف عن شيء يتعلق بسمة أو أكثر من سمات شخصيته. غير أنه ليس ضروريًا أن تتطلب فقرات استبيانات الشخصية معرفة الفرد لنفسه، ففقرات مثل «أذهب إلى المسجد بانتظام» و«أميل إلى قراءة المجلات العلمية» تتعلق بحقائق عن الفرد يسهل عليه ملاحظتها، ومع هذا ربما لا يكون الفرد صادقًا في استجاباته لهذه الفقرات. فالاتجاهات العقلية تؤثر تأثيرًا لا يمكن إغفاله في استجابات الأفراد لفقرات هذه الاستبيانات.

(٤) **الاتجاهات العقلية في الاستجابات:** يتضح تأثير الاتجاهات العقلية؛ أي تحيُّز الاستجابات أو تزييفها Response Sets بدرجة أكثر حدةً في استبيانات الشخصية عن غيرها من الاختبارات والمقاييس، فعلى الرغم من أن مصمِّمي الاستبيانات يُعنون عناية كبيرة بصياغة فقراتها صياغة دقيقة وواضحة لا تحتمل سوء التفسير ، وبحيث

تقيس جوانب يمكن ملاحظتها وتتعلق بحقائق عن حياة الفرد، إلا أن مشكلة تزييف المختبرين لاستجاباتهم تُعد من المشكلات الأساسية فى استبيانات الشخصية. وقد اهتمت كثير من الدراسات بتعرف الدوافع التى تجعل الأفراد يلجأون لذلك، وعلى الرغم من الأبحاث المستفيضة التى اهتمت بضبط تحيز الاستجابات أو تزييفها، إلا أنها لم تستطع استبعاد هذه المتغيرات الدخيلة من أى استبيان للشخصية.

فالفرد ربما يلجأ إلى تزييف استجاباته ليظهر بمظهر الشخص السوى لكى لا يتم حجزه فى مستشفى الأمراض النفسية مثلا، أو ليقنع المسئول عن العمل بأنه شخص مناسب لوظيفة معينة، وعلى العكس ربما يلجأ الفرد إلى تزييف استجاباته ليظهر بمظهر الشخص غير السوى للتهرب من الجندية، أو من جريمة ارتكبها، أو ليحظى بمزيد من التعاطف من جانب أفراد أسرته وأصدقائه، ومع هذا فإن تزييف الاستجابات ربما يكون غير متعمّد أى لا شعورى. لذلك لعلنا لاحظنا فيما سبق أن بعض الاستبيانات حاولت مراعاة هذه المتغيرات الدخيلة، وذلك بتصميم مقاييس فرعية تساعد فى كشف تأثيرها، مثل مقياس (ك) فى استبيان مينيسوتا المتعدد الأوجه (MMPI) الذى يساعد فى تقييم النزعات الشعورية وغير الشعورية للأفراد الذين يلجأون إلى إظهار أنفسهم بصورة معينة، وبالتالى يمكن ضبط أو تحييد أثر هذه النزعات، وكذلك يساعد فى زيادة فاعلية الاستبيان فى التمييز بين الأفراد العاديين وغير العاديين من مرضى المستشفيات النفسية.

وقد سبق أن ناقشنا بالتفصيل الأنماط المختلفة لهذه الاتجاهات العقلية، مثل مراعاة القبول الاجتماعى، والإذعان، وتطرُّف الاستجابات، وغيرها فى الفصلين السابقين. ولعل البحث عن أشكال من الاستبيانات تختلف عن تلك المتوافرة فى وقتنا الحاضر ربما يسهم فى التغلب بدرجة أفضل فى استبعاد أثر هذه الاتجاهات العقلية التى تؤثر فى صدق تفسير الدرجات.

فقد تبين لنا أن الاستبيانات التى تشتمل على فقرات جبرية الاختيار - Forced Choice لم تتغلب على هذه المشكلة، بالإضافة إلى صعوبات تفسير الدرجات التى تستند إلى إطار مرجعى فردى Ipsative Scores. وقد اقترح جولد برج Goldberg شكلا جديداً من أشكال الفقرات التى تتطلب من المُختبر ما إذا كان قد قام بنشاط معين فى الماضى (مثل تسلق الجبال، أو التأثير على المعلمين) مثلا؟ وما رأيه فى هذا النشاط (هل يحبه أم لا يحبه)؟ وما احتمال قيامه بهذا النشاط مستقبلاً؟ ويرى أن هذا الشكل من الفقرات يزودنا ببيانات متنوعة باستخدام مثير أو عبارة واحدة، كما يسمح بتحليل أنماط الاستجابات مثل، «قمت بهذا النشاط، وأحبه، ولكننى لن أقوم به فى

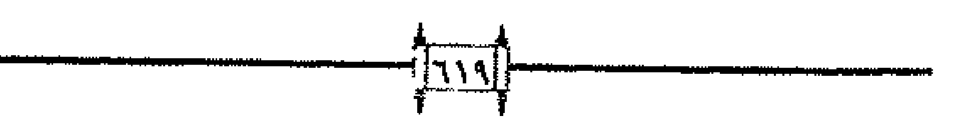
المستقبل» أو «قمت بهذا النشاط، ولم أحبه ، ولكنني أود أن أقوم به في المستقبل»، وغير ذلك من أنماط الاستجابات.

(٥) زمن تطبيق الاستبيانات: تُعد الشخصية من الخصائص المركبة من مكونات متعددة يتطلب قياسها عدة مقاييس فرعية يشتمل عليها الاستبيان، ولكي تتميز الدرجات بالثبات ينبغي أن يشتمل كلا من هذه المقاييس الفرعية على عدد مناسب من الفقرات المتعلقة به لكي تُمثل النطاق السلوكي للسمة المراد قياسها.

لذلك نلاحظ أن استبيانات الشخصية تشتمل على عدد من الفقرات أكبر بدرجة واضحة من عدد الفقرات التي تشتمل عليها معظم المقاييس الجماعية للذكاء أو الاستعدادات الخاصة. فاستبيان جيلفورد وزمرمان يشتمل على ٣٠٠ فقرة، واستبيان كاتل يشتمل على ١٨٧ فقرة، واستبيان مينيسوتا يشتمل على ٥٦٦ فقرة، واستبيان إدواردز يشتمل على ٢٢٥ فقرة، مما يتطلب وقتًا كبيراً في قراءتها، وبخاصة إذا كان الفرد المستجيب ضعيفًا في القراءة، بالإضافة إلى الزمن اللازم لتفسير ما تعنيه كل فقرة من فقرات الاستبيان كما ذكرنا في البند الأول، وبخاصة إذا استند بناء الفقرات إلى أساس إمبيريقي حيث لا ترتبط الفقرات ارتباطًا مباشراً بالسمة المراد قياسها. وتزداد حدة هذه المشكلة إذا تحرّى الفرد الصدق في استجاباته، ونظر إلى نفسه نظرة تحليلية لكي يختار الاستجابة التي تنطبق عليه أو تصف سلوكه.

(٦) تعدُد أسباب أنماط الاستجابات المتماثلة: يجيب بعض الأفراد عن فقرة معينة من فقرات الاستبيان بـ «نعم» لأسباب ربما تكون مختلفة، فإذا استجاب فردان بـ «نعم» لفقرة مثل: «أحب العمل الجماعي» فإن الأول ربما يشعر بالأمان بين أفراد الجماعة، بينما يود الآخر تكوين صداقات، وبذلك يمكن لفردين بنية شخصيتهما مختلفة أن يستجيبا استجابة واحدة لفقرة معينة، وهذا يجعل من الصعب التمييز بين أنماط الشخصية المختلفة.

(٧) تجمُعات السمات: يتطلب تفسير استجابات الفرد أن يكون هنالك تجُمع عام من السمات التي تقيسها فقرات الاستبيان. وعلى الرغم من محاولات جيلفورد، وكاتل، وأيزنك وغيرهم من المشتغلين بالتحليل العاملي تحديد أبعاد أو عوامل للشخصية، إلا أن هذه المحاولات لم تتوصل إلى اتفاق حول هذه العوامل، على عكس النتائج المثمرة التي أمكن التوصل إليها في مجال القدرات العقلية، والميول، والاتجاهات. وربما يرجع ذلك إلى أن أبعاد أو عوامل الشخصية تتفاعل مع المواقف المختلفة بطرق متباينة، فالفرد المسيطر في العمل ربما يكون منقاداً في المنزل، أو مع أصدقائه



المقرَّبين، فأنماط السمات الاجتماعية لدى الأفراد تعتمد على خصائص الموقف الذى يواجهه الفرد.

(٨) تعددُ أساليب بناء استبيانات الشخصية: اتضح لنا مما سبق تعددُ الأساليب المختلفة التي يستند إليها بناء استبيانات الشخصية. ولعل هذا يدعو إلى التساؤل حول أكثر هذه الأساليب فاعلية في هذا الشأن. فهل الأساليب النظرية المنطقية ذات جدوى؟ وهل الأساليب الإحصائية المتعلقة بالتحليل العاملى أكثر فاعلية من الأساليب الإمبيريقية؟

والحقيقة أن إجابة هذا التساؤل تعتمد على معرفة الغرض الذى يُبنى من أجله استبيان الشخصية لكى يكون أكثر فاعلية. فمما لا يدعو إلى الشك أن الاستبيانات التى تستند فى بنائها على الأساليب الإمبيريقية، مثل استبيان الشخصية متعدد الأوجه (MMPI) تكون أكثر استخدامًا في المجالات التطبيقية، وتهتم كثير من الدراسات والبحوث السيكلوجية بإجراء دراسات حولها، وهذه النتيجة تظل قائمة على الرغم من أن الاستبيانات التى تستند فى بنائها إلى أسلوب التحليل العاملى ينبغى أن تتميز نظريا على الأقل بمزايا الاستبيانات التى تستند إلى الأساليب الإمبريقية، بالإضافة إلى أنها يمكن أن تُزودنا بدرجات يسهل تفسيرها (Pervin, 1975).

وربما يرجع سبب عدم تحقق ذلك إلى أن الاستبيانات المتوافرة التى تستند إلى التحليل العاملى لا تقيس جميع أبعاد أو عوامل الشخصية. وبالطبع فإن أداة القياس المثالية يجب أن تشتمل على فقرات يتم انتقاؤها استنادًا إلى الأساليب الثلاثة.

(٩) ثبات درجات استبيانات الشخصية: هناك العديد من الأدلة التى تُبيِّن أن ثبات درجات استبيانات الشخصية، سواء معامل الاستقرار أو التكافؤ أو الاتساق الداخلى، أقل بعامة من ثبات درجات غيرها من المقاييس، وبخاصة مقاييس الجوانب المعرفية. ولكن هل هذا الانخفاض يرجع إلى طبيعة السمات المقاسة أم إلى عملية القياس ذاتها؟ فالشخصية بعامة لا تُعد خاصة مستقرة بدرجة مماثلة لبعض الخصائص الأخرى. فمزاج الفرد يختلف باختلاف المواقف ومن يوم إلى آخر، وكذلك يميل السلوك إلى التغير من ظرف إلى آخر، كما أن خصائص استبيانات الشخصية تُسهم فى خفض ثبات درجاتها.

فبعض المشكلات التى أشرنا إليها فى البنود السابقة، مثل عمومية صياغة فقرات الاستبيان، واختلاف تفسيرانها من فرد إلى آخر تجعل استجابات الأفراد غير متسقة مما يؤثر فى ثبات الدرجات.

كما أن ثبات أو اتساق الصفحات النفسية (البروفيل النفسي) الذى تُزودنا به بعض هذه الاستبيانات يعتمد على عدد فقرات كل مقياس فرعي، حيث إننا نهتم باتساق درجات كل من هذه المقاييس على حدة. فالمقاييس الفرعية القصيرة تؤدى إلى قيم منخفضة لمعامل الثبات.

ونظرًا لأننا نتوقع تغيُّرات معينة في الشخصية نتيجة الخبرة، وتقلُّب الحالة المزاجية، فإننا يجب أن نتوقع بالتالى انخفاض قيم معامل استقرار الدرجات عبر الزمن. ويمكن عندئذ استخدام طريقة التجزئة النصفية أو طريقة كيودر ـ ريتشاردسون التى تتطلب كل منهما تطبيق الاستبيان مرة واحدة لتقدير قيمة معامل اتساق الدرجات، ولكن ينبغى أن تكون السمات أحادية البُعد. وقد أوضحت الأدلة أن استبيان الشخصية الذى يُعتنى ببنائه يمكن أن يحقِّق درجة مرتفعة من الاتساق الداخلى.

(١٠) صدق استبيانات الشخصية: على الرغم من أن الثبات يُعد من مشكلات استبيانات الشخصية، إلا أن الصدق يعد مشكلة أكثر تعقيدًا، فكثير من هذه الاستبيانات تفتقر إلى دراسات تتعلق بجوانب الصدق المختلفة، فيما عدا مؤشرات عن صدقها الظاهرى، أى تعلُّق الفقرات بالسمة المراد قياسها. غير أن هذه المؤشرات ليست فقط غير كافية للتحقق من صدق الاستبيان، بل إنها ربما تؤثر فى الصدق تأثيرًا سلبيًا، فكلما بدت الفقرات مماثلة للسمة المراد قياسها أدى ذلك إلى تزييف المختبرين لاستجاباتهم (Nunnally, 1970) . وقد سبق أن أوضحنا أنه يمكن التحقق من الصدق التنبؤى لمقاييس الجوانب المعرفية، وذلك باختيار محكات خارجية مناسبة، غير أنه يصعب التوصُّل إلى مثل هذه المحكات التى يمكن التنبؤ بها باستخدام استبيانات الشخصية.

فمثلا ما الذى نستطيع التنبؤ به من سمة أو عامل «البساطة» مقابل «التعقد» فى استبيان كاتل، أو عامل «الموضوعية» أو «علاقات المودة» فى استبيانات جيلفورد وزمرمان ، فدرجات مثل هذه الاستبيانات يمكن أن ترتبط بمتغيرات أو محكات أيًّا كانت. غير أنه إذا استُخدمت الأساليب الإمبيريقية فى بناء الاستبيانات، فإن تفسير الدرجات لا يُصبح مشكلة أساسية. فنحن لا نهتم فى هذه الحالة بأسباب اختلاف استجابات مرضى العيادات النفسية عن استجابات الأفراد العاديين مثلا، وذلك لأننا

نلاحظ فقط أن هناك فروقًا ذات دلالة إحصائية بين استجابات كل من المجموعتين، أما أسباب اختلاف الاستجابات فهذا من مهام الأخصائي الكلينيكي، ولكنه لا يهم أخصائي القياس النفسي الذي يحاول التحقق من صدق الاستبيان.

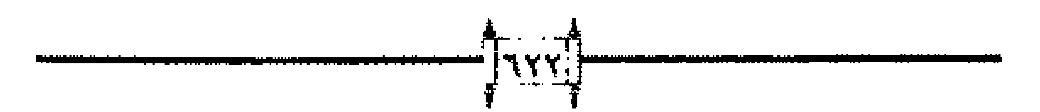
وقد استندت بعض استبيانات الشخصية في التحقق من صدقها إلى الأساليب الارتباطية، وذلك بإيجاد الارتباط بين درجات الاستبيان ودرجات استبيان آخر مماثل له يتميز بالصدق، كما أن بعضًا آخر استند في ذلك إلى أسلوب التحليل العاملي. فقد رأينا فيما سبق أن استبيان العوامل الستة عشر لكاتل استند إلى أحد أساليب التحليل العاملي، بينما استند استبيان جيلفورد وزمرمان إلى نتائج التحليل العاملي.

غير أنه في معظم الأحيان يكون صدق التكوين الفرضي أكثر ملائمة لتقدير صدق استبيانات الشخصية، فسمات الشخصية تُعد بمثابة تكوينات فرضية، والتحقق من صدق هذه التكوينات يتطلب جمع بيانات من مصادر متعددة للتوصل إلى أدلة كافية في هذا الشأن كما سبق أن أوضحنا في الفصل الخامس، فصدق التكوين الفرضي يصعب التحقق منه في وقت قصير أو التعبير عنه بدليل واحد. لذلك ينبغي على مؤلف الاستبيان أن يُقدِّم بيانات متعددة تُلقي الضوء على صدق التكوين الفرضي للاستبيان مسترشدًا في ذلك بمدخل السمات المتعددة والطرق المتعددة Multitrait - Multimethod Approach الذي اقترحه كامبل ، وفيسك Campbell & Fiske وأوضحناه في الفصل الخامس.

فكلما تعددت أدلة الصدق وتنوعت الدراسات المتعلقة به، زادت الثقة في استبيان الشخصية وفي استخداماته المناسبة.

ثانيًا مقاييس غير محددة البنية وغامضة الهدف (الأساليب الإسقاطية):

تُعد الأساليب الإسقاطية Projective Techniques اتجاهًا آخر في تقييم الشخصية يختلف عن الاتجاه السيكومتري الذي تناولناه في الجزء السابق المتعلق باستبيانات أو قوائم الشخصية . وقبل أن نناقش هذه الأساليب ينبغي أن نوضح مفهوم «الإسقاط Projection» الذي تستند إليه. فقد قدَّم فرويد Freud هذا المفهوم في الأدبيات السيكلوجية، حيث عرَّفه بأنه العملية التي ينسب بها الفرد دوافعه ومشاعره وعواطفه إلى أفراد آخرين أو إلى أشياء في العالم الخارجي كدفاع يحول بينه وبين



شعـوره أو إدراكه لخصائصه التي تُهدده، فاستخدام الفرد لعملية الإسقاط تسمح له بإنكار ما بداخله من دوافع غير مرغوبة، وبإسقاط هذه الدوافع لا تُصبح مُدركة كجزء منه وإنما تُصبح جزءًا من الآخرين (Freud, 1911). أما مفهوم «الأسلوب الإسقاطي Projective Technique» ، فقد أصبح معروفًا عقب مقالة فرانك (Frank, 1939) الذي أوضح فيها أهمية أدوات قياس الشخصية التي تتميز بالغموض أو التنكر Disguised Devices، وعارض استخدام استبيانات الشخصية بحجة أنها تستخدم في تصنيف الأفراد، ولا تُعطينا معلومات مفيدة تتعلق بالفرد كإنسان ، وحث على استخدام اختبارات تساعد في فهم واضح للمشاعر والأفكار الخاصة لدى الفرد.

وعندما يُطبق مفهوم الإسقاط على الأساليب الإسقاطية فإنه يُستخدم بالمعنى العام الذي يشير إلى العملية التي تؤثر بها خصائص الفرد، أي بنية شخصيته، في أساليب إدراكه وتنظيمه وتفسيره للبيئة وخبراته، وهذه التأثيرات يمكن أن تتضح وتقاس بدرجة أفضل عندما يواجه الفرد مواقف أو مثيرات غامضة. وقد استخلص ليندزي (Lindzey, 1961) من مراجعته الشاملة لمفاهيم أساليب الإسقاط أن ما يُميّز هذه الأساليب بدرجة أساسية هو حساسيتها للجوانب اللاشعورية أو الكامنة في الشخصية، وتنوع الاستجابات التي يُقدمها الفرد، وتعدُّد أبعاد الشخصية التي يتم قياسها في آن واحد، وعدم إدراك الفرد للغرض من الاختبار، وثراء البيانات والمعلومات المتعلقة بالاستجابات.

ولعل صعوبة التمييز بين مفهوم الإسقاط كما استخدمه فرويد وأطلق عليه الإسقاط الدفاعي أو غير السوي، والمفهوم كما استُخدم في المواقف الاختبارية يرجع إلى التشابه اللغوي بين المفهومين وتداخل المعنى في استخدام كل منهما. فالإسقاط من وجهة نظر فرويد يعتقد فيه الفرد بالفعل أن الصفات أو الخصائص التي أسقطها ليست صفاته أو خصائصه، ولا يكون مُدركا أو واعيا بذلك، وهدفه التخلص من رغباته وأفكاره ودوافعه غير المرغوبة أو المقبولة، أما الإسقاط في إطار المواقف الاختبارية فهو العملية التي يستجيب بها الفرد عندما يقدَّم له عدد من المثيرات الغامضة استجابة نشطة وواعية. ولعل أفضل تعريف للأسلوب الإسقاطي هو ما اقترحه ليندزي (Lindzey, 1961) بأنه أداة تتميز بالحساسية لمظاهر السلوك اللاشعوري أو غير الظاهر، وتسمح بل وتُشجِّع الفرد على أن يقدم استجابات كثيرة متنوعة، أي متعددة الأبعاد، وتستدعي بيانات ثرية تتعلق بالاستجابات ، ولا تتطلب إلا قدرًا ضئيلاً من إدراك الفرد للغرض من الاختبار. ويهتم الأخصائي الكلينيكي الذي يستخدم هذا الأسلوب بالتوصل إلى تفسيرات كلينيكية شاملة من هذه البيانات.

فالأساليب الإسقاطية وما يتعلق بها من اختبارات تسمح للفرد بأن يفرض تكوينا وتنظيمًا خاصا على المثيرات، وبهذا تُعبّر عن تصور دينامي للشخصية.

ويمكن أن تتضح العلاقة بين هذه الأساليب وبين النظريات السيكودينامية في الأسس التي بُنيت عليها الاختبارات الإسقاطية التي سنوضحها بعد قليل. فهذه النظريات تؤكد الجوانب المزاجية الذاتية للفرد، وتعتبر أن السلوك يحدث نتيجة تفاعل قوى متعددة مثل: الدوافع، والحاجات، والحوافز، والصراعات، وأن تكوين الشخصية يتضمن تنظيمات مختلفة مثل الشعور واللاشعور، وأن الشخصية هي العملية التي يفرض فيها الفرد تنظيمًا وتكوينًا على مثيرات خارجية في البيئة المحيطة به. كما تهتم هذه النظرية بالفهم الشامل المتكامل للسلوك من حيث العلاقة بين أجزائه بدلاً من تفسير السلوك بطريقة جزيئية أو على أساس سمات شخصية منفصلة كما في استبيانات الشخصية.

تصنيف الأساليب الإسقاطية :

تتعدد طرق تصنيف الأساليب الإسقاطية، ولعل أبسطها هو التصنيف الثنائي إلى قسمين عامين هما:

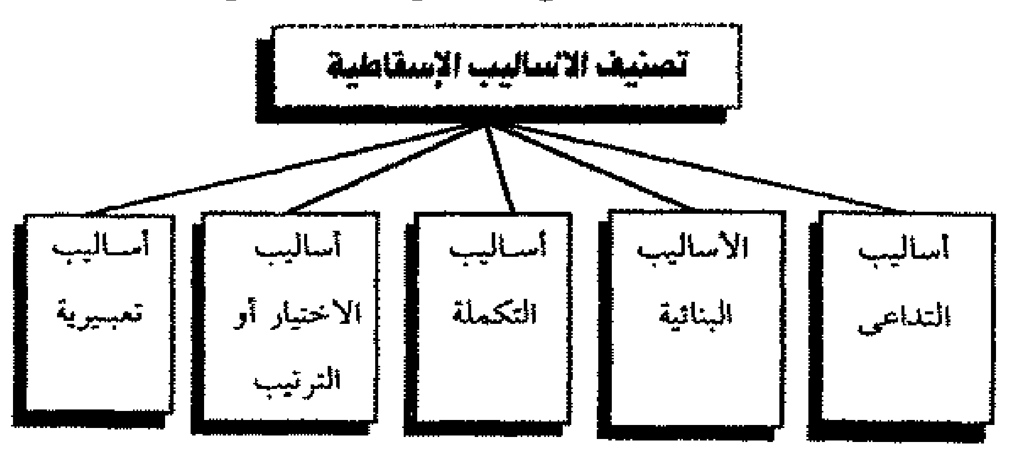
(١) أسلوب يُركز على المحتوى Content - Centered : يهتم هذا الأسلوب بتحليل استجابات الفرد في ضوء الموضوع أو المشاعر التي يعبّر عنها، حيث يُفترض أن هذا المحتوى يكون نتاجًا لتصورات الفرد ويُمثل تخيلاته. فاختبار تفهُّم الموضوع Thematic Apperception Test (TAT) الذي أعده موراي Murray يعد مثالا لهذا الأسلوب الإسقاطي، حيث يتم تحليل الاستجابات في ضوء الموضوعات Themes التي يعبّر عنها الفرد في قصص يكتبها كاستجابة لسلسلة من الصور التي تتميز بالغموض النسبي.

(٢) أسلوب يُركز على الشكل Form - Centered : يستمد هذا الأسلوب أهميته التشخيصية بدرجة أساسية من خصائصه الشكلية، فما يراه الفرد في المثير الغامض يتعلق بعدد من الأقسام الشكلية ويُعد اختبار بقع الحبر لرورشاخ Rorschach Inkblot Test مثالا لهذا الأسلوب، فهذا الاختبار يُصنف استجابات الفرد من حيث بعض الأقسام الشكلية، مثل الحركة، والنزعة إلى استخدام الكل أو الأجزاء في بقعة الحبر، وتنظيم الشكل، والحساسية للألوان والظلال. وتُركز بعض المقاييس الإسقاطية



الأخرى على المحتوى والشكل معاً، مثل اختبار الرسومات، فتحليل الرسومات يستند بدرجة أساسية إلى الأقسام الشكلية، وكذلك على محتوى هذه الرسومات.

غير أن ليندزى (Lindzey , 1959 , 1961) اقترح تصنيفًا أكثر شمولاً وتكاملا من التصنيف السابق للأساليب الإسقاطية المتنوعة فى خمسة أقسام عامة استناداً إلى نمط الاستجابة المطلوبة يوضحه الشكل التخطيطى (١٣ـ٤) التالى:



شكل (١٣ـ٤) يوضح تصنيف ليندزى للأساليب الإسقاطية

(١) أساليب التداعى Association Techniques : وفيها يُطلب من الفرد أن يستجيب بأقصى سرعة لمثيرات مثل الكلمات أو الصور أو المدركات ، ومن أمثلة ذلك اختبارات تداعى الكلمات، وإكمال الجمل، واختبار الرورشاخ، وكذلك جهاز كشف الكذب Polygraph الذى يعتمد على التأثيرات المعلومة للانفعالات الشديدة على ضغط الدم، والنبض ، وعرق راحة اليد.

(٢) الأساليب البنائية Construction Techniques : وفيها يُطلب من الفرد تكوين قصة أو رسم صورة دون أى قيد على استجابته، وتختلف هذه الأساليب عن أساليب التداعى فى أن الفرد فى تكوينه لصيغة فنية يقوم بسلوك معرفى معقد يُوسِّع فيه ويُعدِّل ويُطوِّر تداعياته الأصلية أثناء كتابة القصة أو رسم الصورة. ومن أمثلة ذلك اختبار تفهُّم الموضوع (TAT)، واختبار رسم صورة وكتابة قصة، واختبار بلاكى للصور Blacky Pictures Test .

(٣) أساليب التكملة Completion Techniques :وفيها يُطلب من الفرد أن يكمل مهمة ناقصة بأى طريقة يراها مناسبة، والاستجابة فى هذه الحالة تكون أكثر بطئًا منها فى أساليب التداعى، والمثيرات الأصلية تكون أكثر تعقيداً. ومن ناحية أخرى

تكون نواتج الاستجابات أقل تـعقدًا وأكثر تقييدًا منها فى الأساليـب البنائية، ومن أمثلة ذلك اختبارات تكملة القصص Story Completion Tests.

وأكثر هذه الاخـتبارات شيوعًا قائـمة الجمل الناقصـة التى أعدها روتر Rotter، حيث تشتمل على ٤٠ جملة ناقصة مثل:

«أتمنى لو أننى . . .».

«معظم الناس . . .».

«والدتى . . .».

ويُطلب من الفرد أن يكـمل الجمل بما يختاره دون تحـديد لزمن الإجابة، وهنا يُفتـرض أن الطريقـة التى يكمل بهـا الفرد الجـمل تعكس مـا ينطوى عليـه من دوافع وصراعات وحاجات أو غير ذلك من ديناميات الشخصية. وتُستخدم هذه القائمة وغيرها من الأساليب الـبنائية فى المجـالات الكلينيكية، وكذلـك فى اتخاذ القرارات المتـعلقة بانتقاء الأفراد.

(٤) أساليب الاختيار أو الترتيب Choice or Ordering Techniques : وفيها يُطلب من الفرد أن يختار من بين عدد من الصـور أو البدائل بديلاً أو أكثر وفقًا لمحك معين، مثـل المواءمة أو الفرق أو الاستحـسـان، وغير ذلك، وأحيانًا يُطـلب منه ترتيب البدائل استنـادًا إلى أحد هذه المـحكات، ومن أمثـلة ذلك اختـبار تومكيـنس وهـورن Tomkins - Horn Picture Arrangement Test الذى يشتمل على ثلاثة رسوم عبارة عن خطوط، وتُمـثل أنشطة يقـوم الفرد بترتيـبها لكى تـكوّن قصة، واختـبار زوندوى Szondi Test.

(٥) أساليب تعبيرية Expressive Techniques : وفيها يُطلب من الفرد القيام بدور نشط فى الرسـم أو الطلاء أو اللعب كوسيـلة للتعبيـر عن شخصيتـه، وهنا يكون التركيز على النتاج، وأسلوب أو نمط التعـبير. وكما أوضح ألبورت فإن هذه الأساليب يُفترض أن لها قيمة علاجية وتشخيـصية، حيث إن الفرد هنا لا يكشف عن نفسه فقط، بل يُعبّر عن نفسه أيضًا بطريقة تؤثر فى تكيُّفه الشخصى. ومن أمثلة ذلك مواقف اللعب مع الأطفال، ورسم شخص Draw - A - Person Test، وأساليب الطلاء.

وقد قـدّم ألبورت (Allport, 1933) مفهـوم الأسلوب التعـبيرى Expressive Technique ليشـير إلى استـجابات الفرد التـى يتطلبهـا اختبار الرورشـاخ، وإجراءات اللعب Play Procedures التى تُستخدم عادة فى الأسالـيب العلاجية، وكذلك الدراما



النفسية Psychodrama، والدراما الاجتماعية Sociodrama التى تُستخدم فى المدارس وفى الإرشاد لتحسين العلاقات بين مجموعات من الأفراد.

أمثلة لبعض الاختبارات الإسقاطية:

نظرًا لأن أكثر الاختبارات الإسقاطية شيوعًا واستخدامًا في المجالات الكلينيكية لتقييم الشخصية هما اختبار بقع الحبر لرورشاخ، واختبار تفهم الموضوع لموارى، فسوف نوضحهما بالتفصيل فيما يلى:

اختبار بقع الحبر لرورشاخ:

يُعد هذا الاختبار من الاختبارات الإسقاطية الأكثر شيوعًا، وقد أعده الطبيب النفسى السويسرى هيرمان رورشاخ H.Rorschach أثناء الحرب العالمية الأولى وبعده، ونشر حوله ما يزيد عن أربعة آلاف مقالة وكتابًا وبحثًا (Bernstein & Neitzel, 1980). وعلى الرغم من أن بقع الحبر قد استُخدمت فى الماضى لاستدعاء استجابات من الأفراد فى الدراسات المتعلقة بالتصور العقلى، إلا أن رورشاخ يُعد أول من أدرك بوضوح أهمية استخدام هذه الاستجابات فى تقييم الشخصية. ومع هذا فإن هذا الاختبار أدى إلى انقسام شديد فى الآراء بين علماء النفس، كما سبق أن أشرنا فى عرضنا للإطار التاريخى لقياس الشخصية.

وعلى الرغم من أن رورشاخ كان متأثرًا بنظرية فرويد وما تتضمنه من مفاهيم، مثل اللاشعور وديناميات الشخصية، إلا أنه لم يُصمم الاختبار استنادًا إلى هذه النظرية، غير أن طبيعة بناء الاختبار يبدو أنها قد تأثرت بها بلا شك.

فبينما أكد رورشاخ فى بعض الأحيان الجوانب الإمبيريقية للاختبار، إلا أنه أكد أيضًا دور نظريات علم النفس الدينامى Psychodynamics عند التفكير فى بنائه للاختبار.

ويتكون الاختبار من ١٠ بطاقات متماثلة، بُعدا كل منها ٥,٥ ، ٩,٥ بوصة، خمسة منها ظلالها بيضاء وسوداء، وبطاقتين حمراء ورمادى، وثلاثة بطاقات متعددة الألوان. ويوضح شكل (١٣ـ٥) التالى (Brown, 1970) إحدى بقع الحبر المماثلة لبطاقات الاختبار.



شكل (١٣ـ٥) يوضح إحدى بقع الحبر المماثلة لبطاقات اختبار رورشاخ

وقد صمم رورشاخ المثيرات عن طريق وضع بقع الحبر على ورقة، وقام بطى الورقة بحيث ينتج عنها أشكال متماثلة ولكن غير محددة البنية ، ثم قام بتطبيق هذه البقع (المثيرات) على مجموعات مختلفة من مرضى المستشفيات النفسية، واحتفظ بالبقع التى استدعت استجابات مختلفة من هذه المجموعات ، واستبعد البقع الأخرى.

واستخدم رورشاخ فى محاولاته آلافًا من بقع الحبر حتى استقر فى النهاية على عشر بقع، وصمم إجراءات تقدير الدرجات استنادًا إلى الملاحظات الكلينيكية، وتبلورت هذه الإجراءات بعدما قام بتجريب البقع على مجموعات أخرى من الأفراد العاديين، والمتخلفين عقليا، والفنانين، وغيرهم من الأفراد الذين يتميزون بخصائص معلومة.

وعند تطبيق الاختبار تُعرض كل من هذه البطاقات على المفحوص، ويُطلب منه أن يُوضّح ما يراه فيها أو ما يُمكن أن تُمثله بقعة الحبر، ونظراً لأن بقعة الحبر تُعد مثيراً غامضًا، فإن استجابة الفرد تمثل إسقاطه لمعنى معين على المثير، وهذه المعانى التى تم إسقاطها تُستخدم بدورها كأساس للاستدلال عن بنية شخصيته وديناميتها.

ولا تزال بقع الحبر العشر تُستخدم حتى وقتنا الحاضر، وأُعيد النظر فى نظام تقدير الدرجات، ولكن الجهود التى بذلها رورشاخ فى البداية ظلت أساسية فى معظم

نظم تقدير الدرجات المستخدمة حاليًا. وعند تطبيق البطاقات ينبغي أن يحاول الفاحص جعل المفحوص في حالة عدم توتر، وأن يعمل على تكوين صلة وئام معه، وأن تكون التعليمات كافية فقط بالقدر الذي يُمكنه من استكمال استجاباته. ويتم التطبيق عادة في مرحلتين، في المرحلة الأولى يقدم الفاحص كل بطاقة على حدة للمفحوص، ويطلب منه أن يوضح ما تعنيه البقعة بالنسبة له، ويسجِّل الفاحص الاستجابات والمواضع المختلفة في البطاقة التي ركَّز عليها المفحوص في استجاباته، وما الذي جعل البقعة تأخذ هذا الشكل. وكذلك ملاحظاته لبعض جوانب أداء المفحوص، مثل بطء استجابته الأولى لكل بطاقة ، والزمن الكلي الذي استغرقه في كل بطاقة، وكيفية قلب البطاقة على الوجه الآخر. وفي المرحلة الثانية يقوم المفحوص بفحص البطاقات للمرة الثانية للتعليق على مظاهر بقع الحبر التي جعلته يستجيب استجابة معينة.

ويمكن أن يطلب منه الفاحص أثناء ذلك توضيح بعض الأمور المتعلقة باستجاباته، وبذلك يمكن للفاحص أن يحدِّد الأساس الذي بنى عليه المفحوص استجاباته بدقة أكبر. وأهم ما يميز إجراءات تطبيق هذا الاختبار حرية المفحوص في اختيار ما يراه في بقعة الحبر، وأين يراه، ومحددات إدراكاته.

وتتعدد نظم تقدير الدرجات ، إذ يوجد خمسة نظم متمايزة على الأقل، على الرغم من أن النظام الذي اقترحه كلوبفر Klopfer ، وبك Beck (Exner, 1976) يُعد من النظم شائعة الاستخدام. غير أن هناك اتفاقًا بعامة بين جميع النظم على أن تفسير الدرجات ينبغي أن يستند إلى أبعاد معينة هي : الموضع Location، والمُحدِّدات Determinants، والمحتوى Content . فالموضع يشير إلى الجزء من بقعة الحبر الذي استند إليه المفحوص استنادًا أساسيًا في استجابته، إذ يمكن أن يستخدم المفحوص البقعة ككل، أو تفاصيل عامة، أو تفاصيل بسيطة، أو الظلال البيضاء، ويمكن للفاحص التحقق من ذلك أثناء المرحلة الثانية. وموضع الاستجابات يُعتقد بأنه مؤشر للمدخل العقلي لحياة الفرد من حيث اعتماده على تعميمات مترامية، أو يراعي التفاصيل بانتظام، أو تضيع لديه المعالم في تركيزه الشديد على قضايا جانبية غير ذات أهمية.

والمُحدِّدات تشير إلى مظاهر البقعة التي تُحدِّد الاستجابة، مثل الشكل أم الصيغة، والظلال أم التركيب، واللون ، والحركة. وتقدير درجات المُحدِّدات يعد عملية معقدة وأكثر ذاتية من تقدير الموقع أو المحتوى، فلكل من هذه المظاهر تقديرات معينة ومؤشرات مختلفة.

والمحتوى يشير إلى القسم الذى ينتمى إليه المُدْرَك. ويتم تقدير محتوى الاستجابات فى ثلاثة أقسام تتعلق بالإنسان أو الحيوان أو المُجرّدات كالعدوان أو العاطفة، بحسب مادة المثير الذى يراه المفحوص. وكذلك يمكن تقدير ذلك بحسب شيوع محتوى إدراك الفرد أو أصالته، فتكرار عدد معين من الاستجابات لكل بقعة حبر هو الذى يحدد شيوعها أو أصالتها. ويختلف تفسير شخصية الفرد تبعًا للمحتوى الذى يراه فى معظم البطاقات، فمثلاً يمكن المقارنة بين تفسيرين ربما يتم التوصل إليهما من مجموعتين من الاستجابات ؛ إحداهما رأى فيها المفحوص حيوانات تتصارع، ورأى فى الأخرى أناسًا يشاركون مشاركة كاملة فى عمل ما.

وفى كثير من الحالات يكون محتوى الاستجابات على شكل قصص ويفسرها الفاحص على أنها تعبير عن خصائص شخصية المفحوص وأسلوب حياته. كما يمكن تفسير المحتوى بطريقة رمزية، فمثلاً الانفجار ربما يرمز إلى عدوان شديد، والثعلب يرمز إلى الميل إلى العدوان، والخنزير يرمز إلى الشره أو النهم، والعنكبوت، والأخطبوط، والساحر، والغول، ربما يرمز كل منهما إلى اتجاهات سلبية نحو الأب المسيطر، والنعامة ربما ترمز إلى محاولة الابتعاد عن الصراعات، والأقسام العامة للمحتوى، مثل الطعام والحيوانات التى يؤكل لحمها، والكائنات العدوانية والحيوانات المفترسة والشخصيات التى تمثل القوة والسلطة يمكن تفسيرها رمزيًا تحليليًا (Klopfer & Davidson, 1962) . وينبغى الإشارة إلى أنه ينبغى تقييم كل استجابة استنادًا إلى الأبعاد الثلاثة السابقة، أى الموضع والمُحدِّدات والمحتوى. فالتفسير لا يتم على أساس استجابة واحدة، ولكن تُفسَّر الاستجابة فى علاقتها بالتنظيم الكلى للاستجابات، فعدد الاستجابات التى تنتمى إلى كل من هذه الأبعاد، وكذلك النسب المختلفة التى يتم حسابها منها يحدد تفسير بروتوكول الاختبار ككل. فمثلا الاستجابات المتعددة الجيدة التى تشير إلى الكل ربما تشير إلى الانفعالية، والحركة ربما تشير إلى التخيل، ونسبه عدد استجابات الحركات الإنسانية إلى عدد استجابات اللون ربما ترتبط بدرجة التفكير العقلى أكثر من التوجُّه العملى ، وبطء الاستجابة ربما يفسَّر على أنه قلق، والعدد القليل من استجابات اللون والحركة ربما يُفسَّر على أنه اكتئاب، والاستجابات الكبيرة التى تتعلق بالظلال ربما تشير إلى ضبط الذات (Schafer, 1954) .

لذلك فإن تفسير الاستجابات لهذا الاختبار يتطلب فاحصًا مدربًا تدرُّبًا عاليًا ولديه خبرات كلينيكية واسعة. فالتفسيرات إذن ينبغى أن تعتمد على النمط العام للاستجابات، وكذلك على سلوك الفرد أثناء استجاباته للاختبار، فالفاحص يستخدم ملاحظاته عن

هذا السلوك كمصدر إضافى للبيانات عند تفسيره للاستجابات. فالمفحوص الذى يطلب دائما العون والإرشاد من الفاحص ربما يُفسَّر سلوكه هذا بأنه سلوك شخص غير استقلالى، والمفحوص الذى يبدو متوترًا أو يسأل أسئلة مراوغة ثم ينظر إلى الوجه الآخر للبطاقات ربما يُفسَّر هذا السلوك بأنه سلوك شخص متشكك، وهكذا.

ثبات وصدق اختبار رورشاخ:

إن الذى يطلع على الدراسات والمراجعات التى أشار إليها الكتاب السنوى للقياس العقلى Buros Mental Measurement Yearbook فى طبعاته المتتالية سوف يلاحظ وجهات نظر ناقدة متعددة فيما يتعلق بالخصائص السيكومترية لاختبار رورشاخ. فثبات الدرجات يتباين تباينًا كبيرًا ربما بسبب تنوع وثراء الاستجابات التى تُستخدم فى تقدير معظم خصائص شخصية الفرد، وتباين التفسيرات، مما يؤدى إلى مشكلات تتعلق بتقدير ثبات وصدق الاختبار.

فمعظم طرق تقدير ثبات درجات الاختبارات التى تُبنى استنادًا إلى الأساليب الإحصائية مثل استبيانات الشخصية تناسب هذه الاختبارات لأنها تؤدى إلى بيانات كمية، أما فى الاختبارات الإسقاطية وبخاصة اختبار الرورشاخ، فإننا لا نستطيع الرجوع إلى جدول معايير لمعرفة دلالة أقسام الدرجات كما فى الاختبارات المقننة.

لذلك فإن إحدى طرق تقدير ثبات الرورشاخ هو تحديد مدى اتفاق مجموعة من المُحكمين فى تقدير الدرجات. ولكن هذه الطريقة غير شائعة الاستخدام بسبب اختلاف نظم تقدير درجات هذا الاختبار، وهذا الاختلاف يؤدى إلى عدة مشكلات فى تقدير الثبات بالأساليب المتعارف عليها. فعدم استقرار الدرجات ربما يرجع إلى عدم استقرار طريقة القياس، أو إلى التغيُّر المؤقت الذى يحدث فى الجوانب المزاجية للفرد. كما أن بطاقات الرورشاخ تختلف عن فقرات استبيانات الشخصية أو مقاييس الجوانب المعرفية مما يجعل من الصعب التحقق من الثبات بطريقة التكافؤ. وكذلك تختلف استجابات الفرد لبقع الحبر من بطاقة إلى أخرى، مما يجعل استخدام طريقة التجزئة النصفية غير مناسبة؛ لذلك توصلت دراسات فيرنون (Vernon, 1933)، وهيرتس (Hertz, 1934)، وكرونباك (Cronbach, 1949)، وسوبين (Zubin,1965) التى استخدمت طرقًا مختلفة لتقدير الثبات إلى نتائج متعارضة.

ونظرًا لافتقار الاختبار إلى الثبات، فإن صدقه سوف يكون كذلك، فتقدير صدق اختبار الرورشاخ يُعد أمرًا بالغ التعقيد، إذ يمكن تتبع عدد كبير من الدراسات في هذا الشأن وجميعها أدت إلى نتائج متناقضة، كما لا يوجد اتفاق فيما بينها حول أفضل طرق تقدير صدق هذا الاختبار. وربما تكون إحدى طرق إجراء ذلك هي التحقق من مدى مطابقة التفسيرات الناتجة عن الاختبار بالبيانات المستمدة من مصادر كلينيكية أخرى، ومع هذا فقد تباينت نتائج الدراسات في هذا الشأن.

ويرى بعض الباحثين أن الأساليب المتعارف عليها في تقدير صدق الاختبارات والمقاييس يصعب تطبيقها على اختبار رورشاخ. إذ من الصعب التوصل إلى محكات مناسبة يمكن باستخدامها التحقق من صدق تفسيرات الاختبار للمستويات العميقة أو اللاشعورية للسلوك، كما يصعب تحديد العلاقات بين أجزاء الاختبار والسلوك الفعلي.

لذلك حاول بعض علماء النفس بناء اختبارات بقع حبر أيضًا ولكن تتميز بمزيد من الموضوعية والدقة، ومثال ذلك أسلوب بقع الحبر لهولتزمان Holtzman Inkblot Technique (HIT) الذي يشتمل على صيغتين متكافئتين تتكون كل منهما من ٤٥ بقعة حبر، تتطلب كل منهما استجابة واحدة فقط. وتم انتقاء بقع الحبر استنادًا إلى قيم معامل الثبات بطريقة التجزئة النصفية، وتمييزها بين استجابات الأفراد العاديين والمرضى لكل بقعة منها. ويمكن تفسير درجات كل استجابة في ٢٢ بُعدًا، كثير منها يماثل أبعاد تفسير درجات اختبار رورشاخ، وتشمل الموضع والمُحدِّدات والمحتوى.

ونظرًا لأن هذا الأسلوب أقرب إلى أسلوب بناء استبيانات الشخصية، فإن الثبات في هذه الحالة أفضل من ثبات اختبار رورشاخ (Holtzman, 1968, 1975). غير أنه تبيَّن من مراجعة الأدبيات أنه لا توجد بيانات كافية للمقارنة بين الاختبارين. كما حاول ساندفن (Sandven, 1975) التوصل إلى منهجية جديدة تضم منهجية بناء استبيانات الشخصية، والخصائص الأساسية للأساليب الإسقاطية أطلق عليها القياس الإسقاطي **Projectometry**. غير أن المجال لا يسمح بتوضيح هذه المنهجية وتطبيقاتها المقترحة.

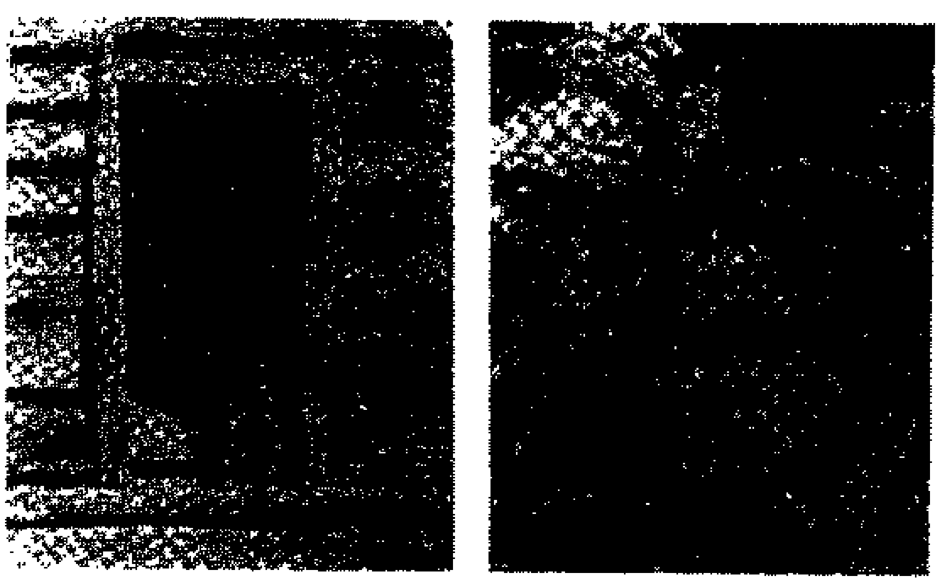
اختبار تفهُّم الموضوع :

أعد اختبار تفهُّم الموضوع The Thematic Apperception Test (TAT) كريستينا مورجان C. Morgan ، وهنرى موراى H. Murray عام ١٩٣٥ بالعيادة النفسية بجامعة هارفارد كأسلوب للكشف عن الأفكار اللاشعورية والخيالات Fantacies . ويختلف هذا الاختبار عن اختبار الرورشاخ فى أن مثيراته أكثر انتظامًا وأقل غموضًا واستجاباته تكون أكثر تحديدًا. فهذا الاختبار يتطلب من الفرد أن يحكى قصصًا تتعلق بصور معينة بحيث يمكن للأخصائيين من ذوى الخبرة تفسير الدوافع الكامنة لدى الفرد، وانفعالاته، وعواطفه، وصراعاته وحاجاته النفسية التى افترضها موراى فى نظريته للشخصية. وهذا الاختبار يلى اختبار رورشاخ فى شيوع استخدامه، وفى عدد الدراسات والبحوث التى استثارها. وقد استمد الاختبار أهميته أولا فى العيادات النفسية، ولكنه أصبح تدريجيًا أداة بحث فى علم النفس الارتقائى، وعلم النفس الاجتماعى، والشخصية، والدراسات الأنثروبولوجية عبر الثقافات ، كما يُستخدم فى تقييم الشخصية فى مجالات الإرشاد وعلم النفس الصناعى.

ويتكون الاختبار من ٣٠ صورة باللونين الأبيض والأسود مرسومة على بطاقات عبارة عن ٤ مجموعات متداخلة تتكون كل منها من ١٩ بطاقة إضافة إلى بطاقة بيضاء، وتختلف كل صورة من حيث محتواها ودرجة غموضها. وتتباين الصور فى درجة انتظامها وبنيتها، إذ تتراوح بين الانتظام التام أو البنية المحددة بدرجة كبيرة إلى البنية غير المحددة على الإطلاق كما فى البطاقة البيضاء.

ويمكن تطبيق الاختبار على الراشدين والمراهقين، وتُعرض البطاقات على الفرد واحدة تلو الأخرى مصحوبة بتعليمات بسيطة، ويُطبق الاختبار عادة على مرتين أو أكثر، وإجراءات التطبيق المعتاد هو أن يُطلب من المفحوص أن يحكى قصة كاملة تتعلق بكل من ١٩ بطاقة يتم اختيارها بحيث تناسب عمره ونوعه، وكذلك البطاقة البيضاء، ويُطلب منه أن يستغرق خمس دقائق فى كل قصة، ويحكى ما يحدث فى الصورة الآن، والأحداث التى أدت إليها. وما سوف تنتهى إليه ، ويجب أيضًا أن تتضمن القصة تفاصيل عن الشخصيات التى فى الصورة، وأنشطتهم وتفكيرهم ومشاعرهم.

ويوضح شكل (١٣ـ٦) إحـدى الصور المماثلة لصور الاخـتبار (Kleinmuntz, 1967).



شكل (١٣ـ٦) يوضح إحدى الصور المماثلة لبطاقات اختبار تفهُّم الموضوع

وهناك أنظمة أو أساليب متعددة أيضًا لتقييم الاستجابات بعضها كمّي والآخر غير كمّي، غير أنه نظرًا لتعقد الأساليب الكمّية وما تتطلبه من وقت طويل، فإن الفاحصين يميلون إلى استخدام الأسـاليب غير الكمّية فى التفسير. ومـعظم هذه الأساليب تأخذ بعين الاعتبار بطل القصة، والحاجـات، والضغوط، والموضوعات، والنواتج. فالبطل Hero هو الشخصية الرئيسة فى كل صورة والتى يتطابق معها المفحوص، حيث تدور القصة حـول بطل واحد يسهل تحـديده من بين الشخصـيات. كذلك يتم التركـيز على خصائص هذا البطل وبخـاصة دوافعه وحاجاته. ومعظم نظم التقـييم بما فى ذلك نظام موراى Murray الأصلى تهتم بشدة واستدامة أو اسـتمرارية وتكرار أو شيوع كل حاجة Need كمؤشرات عن أهمية ومواءمة تلك الحاجة.

أما الضغوط Press فتشـير إلى القُوى البـيئية التى تـتداخل مع أو تُيسّـر تَحقّق الحاجات المختلفة، وهنا أيضًا يكون الاهتـمام مُركزًا على الشدة والاستمرارية والتكرار فى تقييم الأهـمية النسبية لهذه الـقُوى. كذلك تكرار الموضوعات المـختلفة Themes (مثل الاكتئاب)، والنواتج (مثل الفشل) تُعد أيضًا مؤشرات أساسية لهذه الأهمية.

ويمكن أن يُراعى التفسـير طول القصة وتنظيمها وأصالتهـا إلى جانب محتواها، وكذلك نمط الموضوعات، ونـوعية البيئات، ومواقف العلاقات الشخصـية التى يصفها المفحوص.

وعموما فإن كثيراً من الاختصاصيين يستندون في تفسيراتهم لهذا الاختبار على فرضية أن دوافع، ونزعات، وأفعال، ومشاعر، وصراعات بطل القصة أو رمزها الرئيسي تعد جميعها أيضًا من خصائص المفحوص سواء كان واعيًا أو غير واع.

ثبات وصدق اختبار تفهم الموضوع:

إن الخصائص السيكومترية لهذا الاختبار لا تختلف كثيراً عن اختبار رورشاخ، وذلك نظراً لأنه لا توجد إجراءات مقننة لتطبيق الاختبار وتصحيحه وتفسير نتائجه. فإحدى طرق تقدير الثبات في هذا الاختبار تعتمد على اتفاق المحكمين في تحليل كل منهم لبروتوكول الاختبار. ولكي تكون هناك درجة عالية من الاتفاق ينبغي أن يكون المحكمون على درجة عالية من الخبرة والتمرس فيما يتعلق بنظام تقييم الاستجابات للحصول على قيم لمعامل الثبات تقترب من ٠,٩٠ على الأقل لكي تكون هناك ثقة في درجة الاتفاق (Murstein, 1963) . أما الثبات بطريقة إعادة تطبيق الاختبار فإن قيمه تتذبذب بحسب المدة الزمنية الفاصلة بين مرتي التطبيق.

وقد وُجد أن قيمة معامل الاستقرار تقل كثيراً بمرور مدة زمنية تبلغ عدة أشهر؛ لذلك يرى كثير من مستخدمي الاختبار أن معامل الاستقرار لا يناسب هذا الاختبار كما هو الحال في اختبار رورشاخ، وذلك لأننا نتوقع اختلافًا في موضوع القصة ومحتواها الذي يقدّمه الفرد من وقت إلى آخر، وكذلك الحالة النفسية للفرد ربما تتغير بمرور الزمن، واختلاف صيغ القصص ينبغي أن يعكس هذا التغيُّر دون أن يكون هذا دليلا على عدم الثبات. غير أنه ينبغي التركيز على اتساق أقسام للاستجابات أو عدم اتساقها بدلاً من التركيز على المقارنة بين القصص في مرتي التطبيق. أما طريقة التجزئة النصفية في تقدير الثبات فإنها لا تناسب هذا الاختبار كما هو الحال في اختبار رورشاخ. فقد بينت الدراسات أن قيمة معامل الثبات باستخدام هذه الطريقة كانت منخفضة.

ويُعد تقدير صدق هذا الاختبار من الأمور المعقدة أيضًا، فالسؤال الذي يتبادر إلى الذهن هو: هل القصص تعكس حقيقة شخصية الفرد المُختبر أم أنها مجرد ردود فعل نمطية للموقف أي للصور؟ وحتى إذا افترضنا أن القصص تعكس شخصية الفرد، فهل افتراض أن الفرد يتطابق مع البطل، وأن مشكلات البطل والضغوط المؤثرة عليه تعكس حقيقة جوانب حياة المُختبر يُعد افتراضًا صحيحًا؟

ومما يزيد من حدّة المشكلة هو عدم الاتفاق على طريقة واحدة لتقييم الاستجابات ونظام التفسير. ونظرًا لأن الاختبار يُستخدم فى الوصف العام للشخصية، فإن مشكلة تحديد المحك عند تقدير الصدق المرتبط بمحك تُعد من المشكلات المعقدة. غير أن هناك اتفاقًا بين معظم الاختصاصيين فى هذا المجال على أن هذا الاختبار يتميز بصدق محتواه فى تقييم الشخصية.

ومن الجدير بالذكر أن هناك اختبارات أخرى تُعد تعديلات لاختبار تفهُّم الموضوع صُمِّمت بحيث تناسب صغار السن، مثل اختبار القصة المصور لسيموندز Symonds Picture Story Test، واختبار تفهُّم الموضوع للأطفال (CAT)، واختبار الصور لبلاكى Blacky Pictures ،والاختبار السمعى لتفهُّم الموضوع Auditory Apperception Test (AAT)، واختبار أيوا لتفسير الصور Iowa Picture Interpretation Test (IPTT) . كما أن هناك تعديلات أخرى للاختبار لقياس الاتجاهات نحو المدرسة ،والسلطة، والعمل، وغير ذلك تُستخدم أيضًا فى مجالات الإرشاد المهنى والمشروعات البحثية المختلفة.

وكذلك أعد ريتزل ومعاونوه (Ritzel et al. , 1980) اختبارًا بديلاً لاختبار تفهُّم الموضوع، وذلك بانتقاء صور تسمح بتوازن القصص الإيجابية والسلبية، وكذلك بأنشطة متنوعة ومستويات مختلفة من الطاقة لبطل القصة، وذلك بهدف تطوير منهجية اختبار تفهُّم الموضوع بحيث يمكن تقييم الاستجابات تقييمًا كميًا.

مقارنة بين اختبار تفهم الموضوع واختبار رورشاخ:

يُعد كل من اختبار تفهُّم الموضوع واختبار رورشاخ أداة كلينيكية تتطلب اخصائيا متمرسًا لاستخدامها ولتقييم استجابات المفحوصين.

ويفيد اختبار رورشاخ فى الكشف عن الاضطرابات فى عمليات التفكير وبخاصة عند المرضى الذين يتعرضون لاضطرابات ذهانية، بينما يهتم اختبار تفهم الموضوع بالتوافق الاجتماعى. وعلى الرغم من اختلاف الاختبارين فيما يقيسه كل منهما، إلا أن لكل منهما ميزاته وعيوبه. فاختبار تفهم الموضوع يهتم اهتمامًا رئيسًا بمحتوى القصص وخصائص البطل الذى يتطابق معه المفحوص، ويختلف تفسير الاستجابات باختلاف هذه الخصائص. أما محتوى الاستجابات فليس له نفس القدر من الأهمية فى اختبار رورشاخ، فالجزء من بقعة الحبر الذى تشير إليه استجابات المفحوص لا يكون مهمًا بقدر أهمية محددات المدرك (صيغة أو لون منطقة معينة من البقعة).

ونظراً لأن هناك العديد من الأدلة التى أوضحت افتقار كل من الاختبارين إلى الخصائص السيكومترية مثل الثبات والصدق، وبخاصة اختبار رورشاخ، فإن كثيراً من علماء النفس الكلينيكى يُفضِّلون استخدام الاختبارين معًا فى التشخيص والتوصل إلى فهم متكامل للشخصية.

تقييم الأساليب الإسقاطية:

اتضح من عرضنا السابق تنوع الأساليب الإسقاطية واختلاف أدواتها وما تشتمل عليه من مثيرات، وما تتطلبه من استجابات، وما تتوصل إليه من تفسيرات. فالمثيرات التى تستدعى الاستجابات الإسقاطية يمكن أن تكون كلمات أو جملا أو قصصًا ناقصة، أو بقع حبر، أو صوراً محددة أو غير محددة البنية، أو عرائس ولعب، أو رسومًا أو لوحات ملونة، أو غير ذلك. وقد صنفها ليندزى Lindzey كما رأينا فى خمسة أقسام عامة. ونظرًا لأن الأساليب المختلفة تكشف عن أوجه متباينة إلى حد ما للشخصية على الرغم من تداخل هذه الأوجه، فإنه يفضل استخدام أكثر من أسلوب فى عملية التقييم.

ميزات الأساليب الإسقاطية:

تتميز الأساليب الإسقاطية بأن مثيراتها وتعليماتها غير محددة البنية Unstructured أو غامضة مما يسمح بحرية الاستجابة وتنوعها، وتتباين الأساليب فى درجة غموض مثيراتها. فبعضها مثل اختبار رورشاخ تكون مثيراته أكثر غموضًا من اختبار تفهم الموضوع أو إكمال الجمل مما يؤدى إلى تباين ملحوظ فى مدى الاستجابات. ومعظم هذه الأساليب تكون غامضة الهدف أيضًا، مما يجعل من الصعب على الفرد المختبر تزييف استجاباته، وذلك لأن طرق تقييم الاستجابات أو تقدير الدرجات وتفسير الاستجابات يكون غير مألوف له. وكذلك تسمح هذه الأساليب بحرية الأخصائى الكلينيكى فى سَبْر وتوضيح ما يقصده المُختبر. كما أن معظمها لا يحتاج إلى مهارة فى القراءة، ويمكن أن تُطبَّق على الأميين والأطفال الصغار، وتتميز أيضًا بعدم وجود استجابة صحيحة وأخرى خطأ، أو استجابات محددة مسبقًا، حيث يتم تفسير الاستجابات كيفيا أو تصنيفيا بطرق مختلفة. فبعضها مثل اختبار رورشاخ يركز التفسير على الخصائص الشكلية، وفى غيره مثل اختبار تفهم الموضوع أو إكمال الجمل يكون التركيز على المحتوى. وتستند معظم الأساليب على فرضية أن إضفاء

الفرد بنية على المثيرات الغامضة أو ردود فعله لهذه المثيرات يعكس نزعاته وتصوراته الشخصية وبخاصة نزعاته اللاشعورية وميكانيزماته غير اللفظية.

عيوب الأساليب الإسقاطية:

من أهم عيوب هذه الأساليب افتقارها إلى التقنين Standaradization، فالاختبارات الإسقاطية ليست اختبارات بالمعنى السيكومترى الدقيق، وإنما أدوات أو أساليب يستخدمها أخصائى مدرب تدريبًا عاليا فى تقييم الشخصية والكشف عن جوانبها. وعلى الرغم من أن هناك جهودًا لتقنين عرض المثيرات، إلا أنه تظل هناك فروق فى المداخل التى يستخدمها الفاحصون المختلفون. كما أن نتائج التقييم ووصف شخصية المختبر تعتمد على الأحكام المستبصرة للفاحص.

فالفاحص لا تقتصر مهمته على تطبيق الاختبار، وإنما يُعد جزءًا من الأسلوب الإسقاطى ذاته والمواد المستخدمة؛ لذلك يتباين الفاحصون فى مهارتهم وفاعليتهم فى التوصل إلى مواصفات شخصية المُختبرين، وينعكس ذلك على صدق الأسلوب المستخدم. فتقييم الشخصية باستخدام الأساليب الإسقاطية يُعد فنًّا أكثر من كونه علمًا. ومع هذا فإن جهودًا قد بُذلت لتقنين هذه الأساليب، وبخاصة اختبار رورشاخ من حيث بناء صيغ من الاختبار تعتمد على الاختيار من متعدد بدلاً من الاستجابات غير المقيدة. ففى هذه الحالة تُقدم مع كل بقعة حبر قائمة باستجابات يختار الفرد من بينها الاستجابة القريبة مما يراه فى البقعة. وبدائل الاستجابات تم انتقاؤها من سجلات الأفراد العاديين وتقارير غير العاديين. كما أن أسلوب بقع الحبر لهولتزمان Holtzman الذى سبق أن أشرنا إليه يُعد من الأساليب الإسقاطية التى تماثل استبيانات الشخصية فى الأساس الفكرى الذى استند إليه وفى بنائه، وبذلك يُعد أكثر تقنينًا من اختبار رورشاخ.

ومن بين أوجه قصور الأساليب الإسقاطية افتقارها أيضًا إلى المعايير Norms، أو استنادها إلى معايير مقتضبة أو مستندة إلى جماعات مرجعية غير محددة تحديدًا دقيقًا. وعدم توافر هذه المعايير يؤدى إلى تفسير الاستجابات فى ضوء الخبرة الكلينيكية، مما يسمح بتدخل كثير من الجوانب الذاتية والتحيز فى عملية التفسير. وقد أسهمت هذه العوامل مجتمعة فى القيم المنخفضة لمعامل الثبات وعدم وجود أدلة

كافية تتعلق بالجوانب المختلفة لصدق هذه الأساليب وجدوى استخدامها كما سبق أن أوضحنا.

وعلاوة على ذلك فإن جميع الأساليب الإسقاطية تتطلب التطبيق الفردى مما يجعل عملية التطبيق وتقييم الاستجابات وتفسيرها مكلفا. كما يجب أن يقوم بتطبيق هذه الأساليب، وبخاصة اختبار رورشاخ،واختبار تفهم الموضوع،والدراما النفسية والاجتماعية، واللعب،واختبارات الرسم،أخصائيون تلقوا تدريبًا خاصًا على استخدامها وتفسير نتائجها.

وعلى الرغم مما يشوب هذه الأساليب من عيوب، إلا أنها تزوّد الأخصائيين ببيانات تتعلق بديناميات الشخصية يصعب الحصول عليها باستخدام الاستبيانات أو أى أساليب أخرى. غير أنه ينبغى عدم الاستناد إلى نتائجها فى اتخاذ قرارات مهمة تتعلق بالأفراد فى ضوء الطرق التفسيرية الحالية. كما لا ينبغى استخدامها فى الإرشاد المهنى أو انتقاء الأفراد، ولكن يمكن الإفادة من هذه الأساليب فى الأغراض البحثية من أجل التوصل إلى إجراءات كمية لتقييم الاستجابات لكى تصبح أكثر موضوعية، وكذلك لبحث صدق متغيراتها فى علاقتها باختبارات أخرى تم التحقق من صدقها، والتحقق من صدقها التجريبى فى ضوء مجموعات محكة مختلفة مثل مرضى البارانويا، والفصام. فإذا أجريت مثل هذه البحوث، فإن الأساليب الإسقاطية يمكن أن تصبح واحدة من أفضل الاختبارات السيكلوجية (Kline, 1976).

ثالثا: مقاييس غير محددة البنية وواضحة الهدف (الملاحظة، والمقابلة، ومقاييس مفهوم الذات):

اعتمدت مقاييس الشخصية التى أوضحناها فى الجزء السابق على الاستبيانات أو الاختبارات الإسقاطية التى تتطلب ورقة وقلمًا، وتستدعى استجابات فى مواقف اختبارية اصطناعية. وعلى الرغم من أن هذه المواقف تكون مقننة وتسمح بضبط المتغيرات الدخيلة، إلا أنها تختلف عن مواقف الحياة الفعلية. إذ ربما يستجيب الفرد لهذه المقاييس استجابات لا تتفق ونمط سلوكه المميز فى حياته اليومية؛ لذلك لا نستطيع أن نعتمد على هذه المقاييس وحدها اعتمادًا تامًا فى تقييم متغيرات الشخصية، وإنما ينبغى استخدام أساليب أخرى تتعلق بملاحظة السلوك ملاحظة مباشرة فى مواقف طبيعية، أو إجراء مقابلة وجهًا لوجه. وتؤكد هذه الأساليب وضوح الهدف من القياس، حيث إن إخفاء الهدف ربما يعوق الحصول على المعلومات المتعلقة بخصائص شخصية

الفرد، وتُعدُّ بعض أساليب الملاحظة Observations، والمقابلات المباشرة Direct Interviews، والسيرة الذاتية للفرد كما يُقرِّرها بنفسه Autobiography ، أمثلة للمقاييس غير محددة البنية وواضحة الهدف، حيث تستند إلى المدخل الفينومونولوجي Phenomenological Approach. ويرى أصحاب هذا المدخل أن تحليل الشخصية إلى مجموعة من المكونات أو السمات يتنافى مع التنظيم المتكامل والديناميى للشخصية، كما يرون أن اهتمام نظريات التحليل النفسى بالتنظيم اللاشعورى فى وظائف الشخصية يغفل الجوانب المهمة التى ينبغى التركيز عليها، مثل الإدراك، والمعانى، والمشاعر، والذات. ويؤكدون أنه يمكن النظر إلى الفرد على أنه يستجيب للعالم الخارجى فى ضوء إدراكاته المميزة له، فمجاله الفينومونولوجى يُعد جزءًا من بيئته الفيزيائية التى يُدركها والتى يكون لها معنى خاص بالنسبة له، بينما تُعد «الذات Self» جزءًا من هذا المجال المتعلق به شخصيًا؛ لذلك فإن الفرد يعمل جاهدًا لتحقيق ذاته، أى إحداث تطابق أو تناسق بين ذاته الواقعية وذاته المثالية (Rogers, 1961, 1964).

وسوف نُقدم فيما يلى أمثلة لبعض أساليب تقييم الشخصية التى تستند إلى هذا المدخل، مثل أساليب ملاحظة السلوك، والمقابلة. وتختلف هذه الأساليب فى مدى تحديد بنية المواقف التى تتضمنها.

(أ) أساليب ملاحظة السلوك Observation Techniques:

تلعب أساليب الملاحظة المباشرة لسلوك الأفراد دورًا أساسيًا فى تقييم الشخصية سواء فى الفصل المدرسى، أو مواقف العمل، أو فى مراكز الإرشاد، أو المواقف الكلينيكية. وقد سبق أن أوضحنا فى الفصل العاشر دور الملاحظة والمقابلة فى تقييم الأفراد فى مواقف العمل المختلفة. وسوف نوضح بإيجاز فى هذا الجزء بعض أنماط أساليب الملاحظة سواء فى مواقف واقعية أو محاكاة لهذه المواقف، أى مواقف اصطناعية، وكيفية استخدامها فى جمع معلومات تتعلق بشخصية الفرد، والتنبؤ بسلوكه المستقبلى. وتتميز هذه الأساليب بأنها يمكن من خلالها ملاحظة سلوك الفرد كما يحدث تلقائيًا، وتكون الملاحظة مستقلة عن قدرة الفرد أو رغبته فى تقرير سلوكه.

الملاحظة فى مواقف واقعية، ومواقف مضبوطة:

يمكن الحصول على بيانات تتعلق بسلوك الأفراد عن طريق ملاحظتهم فى مواقف واقعية Natural Observation لا يتحكم فيها القائم بالملاحظة، حيث يُسجِّل



سلوك الفرد أو مجموعة من الأفراد أثناء نشاطهم المعتاد دون أن يحاول التدخل أو وضع قيود أو شروط معينة على موقف معين. فمثلا يمكن أن يقوم المعلم بملاحظة سلوك الأطفال أثناء اللعب فى فناء المدرسة، أو يقوم الأخصائى الكلينيكى بزيارة منزل العميل لملاحظة الأنماط السلوكية لأفراد أسرته. وعلى الرغم من أن تواجد القائم بالملاحظة ربما يؤدى إلى إحداث تغيرات فى سلوك من يلاحظهم ، إلا أن عملية الملاحظة تفيد فى الحصول على معلومات أولية أو انطباعات تتعلق بالفرد.

أما الملاحظة فى مواقف تجريبية مضبوطة Controled Observation فإنها تتم داخل المختبر أو فى مواقف تجريبية يتحكم فيها القائم بالملاحظة، حيث إن إجراء ذلك ميدانيا يكون مكلفًا وغير مناسب ويؤدى إلى نتائج معقدة ومربكة؛ إذ يمكنه تنظيم الموقف وتعديله وإدخال متغيرات تستدعى استجابات معينة وفقًا لاحتياجاته البحثية لكى يُحدد كيفية استجابة الأفراد لهذا الموقف، مثل الملاحظة غير المباشرة للتفاعل الاجتماعى بين مجموعة من الأفراد أثناء أدائهم مهمة معينة يتم تنظيمها تنظيمًا مناسبًا. ففى هذه الحالة ربما يُراقب القائم بالملاحظة هذا التفاعل من خلال مرآة ذات اتجاه واحد بحيث لا يستطيع الأفراد رؤيته، وبذلك لا يؤثر تواجده فى السلوك المعتاد للأفراد. أو ربما يود مشاركتهم فى العمل دون علمهم بأنه يقوم بملاحظتهم، فعندئذ تكون الملاحظة بالمشاركة الفعلية Participant Observation. وفى هذه الحالة يستطيع القائم بالملاحظة تسجيل العديد من مظاهر السلوك الانفعالى والحركى والتفاعلى. وجميعها تُعد من المظاهر غير اللفظية لسلوك الفرد، وتُقدِّم صورة واقعية عن شخصيته بدرجة أفضل من التقرير الذاتى للفرد عن نفسه. كذلك ربما يهتم باحث نفسى بسلوك محدد، مثل استجابات الأفراد لمواقف إحباط معينة ربما يندر حدوثها فى الحياة الواقعية، فعندئذ يمكنه تصميم مواقف تجريبية محكمة تؤدى إلى هذه الاستجابات.

وقد سبق أن أوضحنا فى الفصل العاشر أنواعًا أخرى من المواقف الاصطناعية ، مثل عينات العمل Work Samples، والمحاكاة Simulation ، والاختبارات الموقفية Situational Tests.

مصادر الخطأ فى أساليب الملاحظة :

إن أساليب الملاحظة شأنها شأن غيرها من أساليب قياس الشخصية أو تقييمها تؤثر فيها مصادر خطأ متعددة. ومن بين هذه المصادر ما يلى:

(١) **التحيز:** تعتمد أساليب الملاحظة اعتمادًا أساسيًا على القائم بالملاحظة، والإجراءات المنظمة التى يلتزم بها أثناء عملية الملاحظة، وما تتطلبه من تسجيل دقيق

للسلوك المستهدف. غير أنه من المتعارف عليه أن فردين يُمكنهما إدراك الحدث نفسه إدراكًا مختلفًا، مما يؤدى إلى أخطاء الملاحظة، فالعامل الإنسانى يلعب بلا شك دورا فى ذلك. فمشاعر القائم بالملاحظة تجاه فرد معين أو جماعة معينة أو نشاط ما ربما تؤدى إلى تحيزه إيجابًا أو سلبًا، مما يؤثر فى مصداقية الملاحظة. غير أنه يمكن التغلب على ذلك جزئيًا بتقنين عملية الملاحظة وإجراءاتها، واقتصار القائم بالملاحظة على تسجيل السلوك وفق أقسام محددة مسبقًا.

(٢) القيم الشخصية: يمكن أن تؤثر أيضًا القيم الشخصية للقائم بالملاحظة فيما يراه أو يدونه. إذ ربما يُركز فى ملاحظته على ما يراه متسقًا مع قيمه ويغفل ما عدا ذلك أو العكس. كذلك ربما تلعب الاتجاهات العقلية دورًا فى توجيه الملاحظة وجهة معينة مما يؤدى إلى اختلاف النتائج إذا ما قام فرد آخر بالملاحظة.

(٣) توقعات جماعة معينة: فسلوك جماعة منظمة من الأفراد، مثل أعضاء أحد النوادى، أو إحدى الجمعيات، أو طلاب فصل معين يكون عادة متسقًا ومتوافقًا مع معاييرها وأخلاقياتها مما يجعل القائم بالملاحظة مدركًا لما هو مقبول فى الجماعة، وما هو غير مقبول. وربما تتأثر ملاحظته بهذا الإدراك المتعلق بتوقعاتها لمسايرة الجماعة أو الامتثال لها.

بعض الاقتراحات للتغلب على أخطاء الملاحظة :

على الرغم من الصعوبات المتعلقة بالحصول على بيانات دقيقة ومناسبة للغرض من الملاحظة، إلا أنه يمكن التغلب على هذه الصعوبات وتقليل أخطاء الملاحظة لكى تكون أكثر صدقًا بالاسترشاد بالاقتراحات التالية:

(١) تحديد جوانب السلوك المراد ملاحظته: فقد سبق أن أوضحنا فى كثير من المواضع أن تحديد محتوى المجال المراد قياسه فى ضوء الأهداف المرجوة يُعد خطوة أساسية فى كثير من الاختبارات، وبخاصة الاختبارات التحصيلية. كذلك فإن تحديد جوانب أو أقسام معينة من السلوك الذى نود ملاحظته فى ضوء الغرض من الملاحظة والسمات المرجوة يُعد من الأمور الضرورية.

فإحدى المشكلات التى يواجهها القائم بالملاحظة هو عدم معرفته ما ينبغى ملاحظته، كما أنه لا يستطيع ملاحظة وتسجيل كل ما يحدث فى موقف معين. فلكى تكون الملاحظة منظمة ينبغى تحديد طبيعة البيانات والمعلومات المناسبة التى ينبغى الحصول عليها، وكذلك الأسلوب الذى سوف يُستخدم فى تسجيل عينات ممثلة من

السلوك. ويمكن الاسترشاد فى ذلك بأسلوب معاينة الأحداث Incident Sampling Technique،الذى يتعلق بملاحظة وتسجيل أحداث محددة، أو أسلوب المعاينة الزمنية Time Sampling، الذى يزيد من فاعلية الملاحظة عن طريق إجراء سلسلة متتابعة من الملاحظات يستغرق كل منها دقائق محدودة لمدة يوم واحد أو أكثر. وهذه الأساليب لا يكون لها تأثير فى الاستجابة كرد فعل Unobtrusive Measures، نظراً لأن المشكلة الرئيسة فى قياس الشخصية هى أن عملية القياس ربما تؤثر فى سلوك الفرد المستجيب، وبالتالى تؤثر فى القياسات التى نحصل عليها، أى أن القياسات تتفاعل مع الفرد وتؤدى إلى تغيير نظرته إلى نفسه، أو أن يختار دوراً معيناً لمحاولة تحقيق ما يرى أنه يتفق ومطلب القائم بالملاحظة.

(٢) تدريب القائمين بالملاحظة: فعلى الرغم من التحديد الإجرائى لجوانب السلوك المراد ملاحظته، إلا أنه يظل هناك احتمال عدم اتفاق بين القائمين بالملاحظة لأسباب متعددة. لذلك ينبغى تنظيم جلسات لتدريبهم على كيفية الملاحظة الموضوعية لعينة من السلوك، والتمييز بين الحقائق والتفسيرات فى تسجيل ملاحظاتهم وكتابة التقارير، والمقارنة بين النتائج التى توصل إليها كل منهم ومناقشة الاختلافات فيما بينها من أجل تحقيق الاتساق فى التفسيرات، وتقنين أقسام السلوك المرجو ملاحظته.

وتتطلب أساليب الملاحظة استخدام بعض الأجهزة التى يمكن أن تتراوح بين أجهزة بسيطة وقليلة الكلفة نسبيا، مثل أجهزة التسجيل، وساعات التوقيت، وأجهزة مرتفعة الكلفة، مثل الأجهزة الإلكترونية المعقدة التى يمكن استخدامها فى عرض المثيرات، وتسجيل السلوك، وبخاصة إذا كان الهدف قياس الاستجابات أو ردود الفعل الإدراكية والحركية، مما يتطلب تدريب القائمين بالملاحظة أيضاً على كيفية استخدام هذه الأجهزة استخداماً مناسباً.

(٣) تكميم الملاحظات: فملاحظة بعض مظاهر السلوك، مثل التفاعل الاجتماعى والاستجابات الانفعالية، وسلوك المعلم وإدارته للفصل المدرسى تتطلب التعبير عن هذه الملاحظات بمقادير كمية. وهذه المقادير تُمثِّل عادة تكرار حدوث سلوك معين فى مدة زمنية محددة. غير أنه ينبغى فى هذه الحالات تصميم خطة زمنية دقيقة يحدد فى ضوئها عدد مرات حدوث السلوك الذى نهتم بملاحظته.

(٤) تحديد إجراءات تسجيل السلوك: فالبيانات التى يتم جمعها عن طريق الملاحظة ينبغى أن تتميز بالدقة، وهذا يتطلب تسجيل السلوك فور حدوثه حتى لا يتأثر ذلك بالأخطاء التى تنجم عن العمليات الانتقائية فى التذكر.



ولعل التقنيات المتطورة للملاحظة وتنظيم البيانات سوف تسهم فى التغلب على مشكلات التسجيل الفورى الفاعل للسلوك أو الأحداث التى تجرى ملاحظتها. فمثلا يمكن تطوير نظام ترميزى لأقسام السلوك أو الأحداث المعينة ، كأن يُعطى الرمز (أ) لحدث أو سلوك معين، والرمز (ب) لحدث أو سلوك آخر مرتبط بالظاهرة موضع الملاحظة وهكذا. كما يمكن تحديد مرات تكرار الرمز فى مدة زمنية معينة أثناء عملية الملاحظة.

ونظراً لأن إجراء ذلك ينبغى أن يستند إلى أساليب منهجية أكثر اتساقًا ودقة، فسوف نوضح فيما يلى بعض هذه الأساليب التى تُستخدم فى تسجيل الملاحظات ، وهى موازين تقدير السلوك Behavioral Rating Scales، وقوائم المراجعة Checklists .

موازين تقدير السلوك :

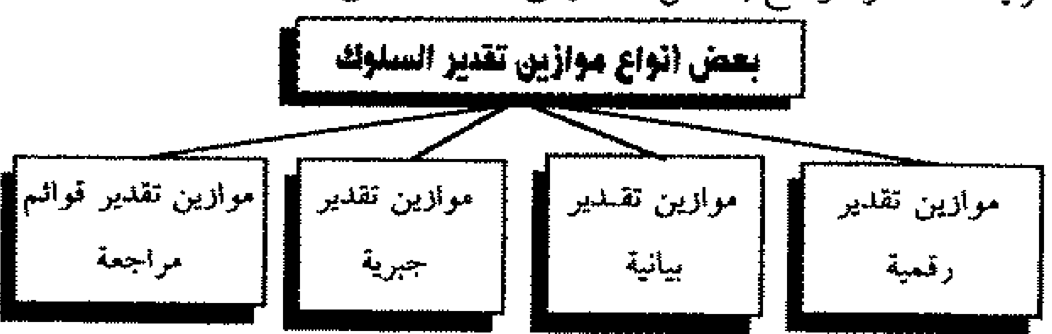
تُعد موازين تقدير السلوك من أكثر الأساليب استخدامًا في تقييم الشخصية، حيث تشتمل على السلوك المرجو ملاحظته فى مواقف واقعية. فالتقدير Rating يتضمن الحكم على خصائص أو سمات الفرد في مواقف مضبوطة ، ويختلف الحكم على السلوك Behavioral Rating عن الحكم على السمات Trait Rating فى أن الأول يتعلق بالتعميم والاستدلال. وميزان تقدير السلوك يُعد أداة تستخدم فى تسجيل قيم تقديرية لما نود قياسه، وبذلك يُفترض أنه يزود القائم بالملاحظة بوسيلة لتقدير السلوك بمزيد من الدقة والموضوعية.

وقد سبق أن قدمنا فى الفصل الثالث عشر أنواعًا متعددة من موازين تقدير الاتجاهات ، وأوضحنا أساليب بنائها وتحليل فقراتها، وكذلك كيفية إعدادها فى الفصل المدرسى. غير أن هذه الموازين تُستخدم فى تقدير الفرد لاتجاهاته بنفسه، أى أنها تُستخدم فى التقدير الذاتى Self-Rating، بينما تُستخدم موازين تقدير السلوك بواسطة القائم بالملاحظة لتقدير سلوك فرد يُفترض أن دوره يكون سلبيًا فى عملية التقدير أو الحكم، ويُعيِّن القائم بالملاحظة قيمة رقمية للسلوك الملاحظ. ومع هذا فإن موازين التقدير لا تختلف فى كل من الحالتين. فميزان تقدير السلوك يُعد أسلوبًا منظمًا للحكم على سلوك ملاحظ وتسجيله وتقريره، وبذلك يمكن التمييز بدرجة أكثر دقة بين الأفراد فى سمة معينة. كما يمكن استخدامها فى تقييم الإجراءات والعمليات والنواتج، والنمو الشخصى والاجتماعى للطلاب فى خصائص متعددة، مثل الالتزام، والتحمس،

ومراعاة الآخرين، والانبساط ، وغير ذلك. وكذلك تُستخدم فى تقييم أداء العاملين وسماتهم الشخصية كما سبق أن أوضحنا فى الفصل العاشر ، وفى تقدير طلاب الجامعات لفاعلية أداء أساتذتهم.

أنواع موازين تقدير السلوك :

تتعدد أنواع موازين تقدير السلوك Behavioral Rating Scales كما سبق أن أوضحنا فى فصول سابقة، غير أننا سوف نتناول مرة أخرى بإيجاز بعض هذه الأنواع الرئيسة كما هو موضح بالشكل التخطيطى (١٣ـ٧) التالى:



شكل (١٣ـ٧) يوضح بعض أنواع موازين تقدير السلوك

موازين تقدير رقمية Numerical Ratings :

تُعد أبسط أنواع الموازين، حيث يُعيِّن القائم بالملاحظة قيمة عددية أو رقما يدل على درجة توافر سلوك أو خاصة معينة لدى الفرد، مثل النشاط العام، أو الجدّية، أو الكفاءة الشخصية،وغير ذلك. ومعظم الخصائص الوجدانية والشخصية يمكن تقييمها بموازين تقدير رقمية. ويُعد أسلوب تمايز معانى المفاهيم Semantic Differential Technique الذى سبق أن أوضحناه بالتفصيل فى الفصل الثانى عشر نوعًا من موازين التقدير الرقمية، ويمكن أيضًا باستخدامه تقييم كثير من هذه الخصائص.

وفى موازين التقدير الرقمية تُقدَّم الخاصة أو السمة فى عبارة معينة يليها قيم تتراوح بين ١ ، ٥ مثلا، ومثال ذلك ما يلى:

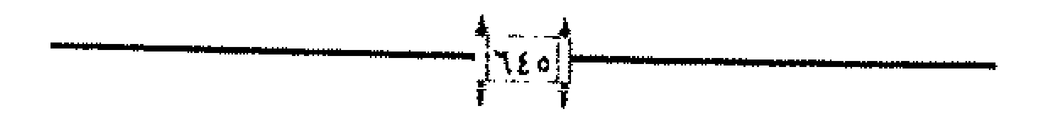
« يتسم بالاكتفاء الذاتى فى علاقته بالآخرين »

١　　٢　　٣　　٤　　٥

حيث ١ = بدرجة مرتفعة ، ٢ = بدرجة فوق المتوسط

٣ = بدرجة متوسطة ، ٤ = بدرجة أقل من المتوسط

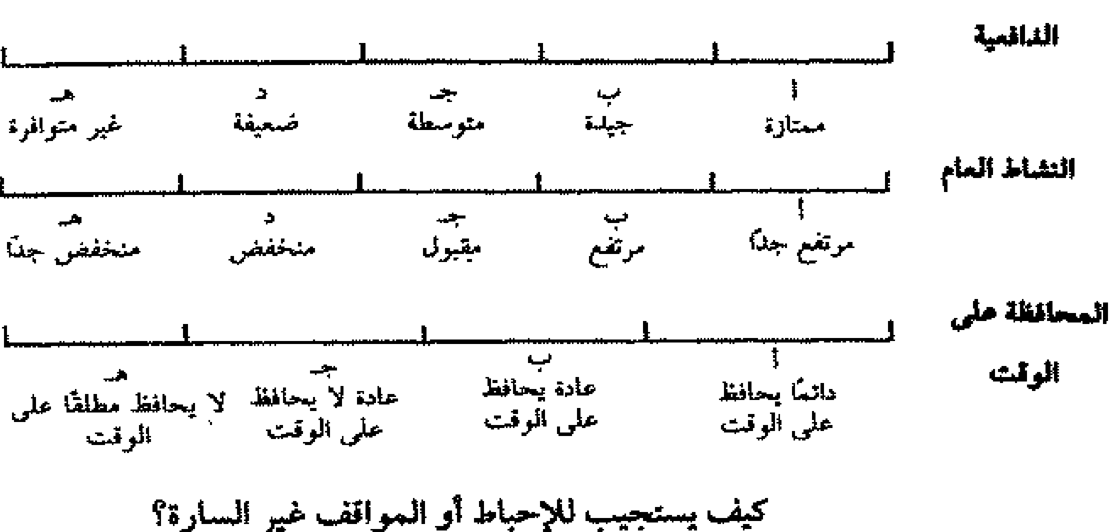
٥ = بدرجة منخفضة



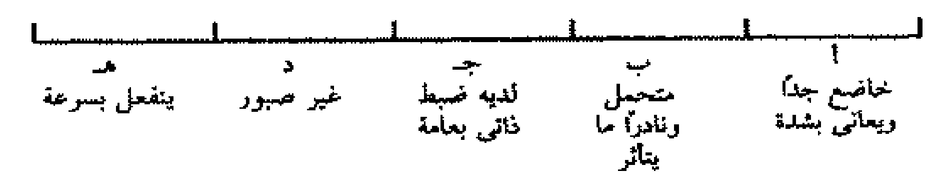
وعادة تكون هذه القيم العددية وما تشير إليه مشتركة بين جميع فقرات ميزان التقدير.

موازين تقدير بيانية Graphic Rating Scales :

يُعد هذا النوع من الموازين أكثر استخدامًا من موازين التقدير الأخرى، ويتطلب أيضًا أن يعيِّن القائم بالملاحظة قيمة عددية للخاصة المعينة، غير أنه في هذه الحالة يكون هناك خط أفقي أو رأسي يُمثِّل خصائص أو أوصاف للسلوك على مُتصل، ويُقسَّم الخط إلى عدد من المسافات تمثِّل عدد التقديرات لكل قسم مُحدد بعنوان معيَّن أسفله. ويستطيع القائم بالملاحظة وضع علامة في أي مكان يراه مناسبًا على الخط دون تقيُّد بنقط معينة . ويوضح شكل (٨ـ١٣) بعض الأمثلة لأنواع موازين التقدير البيانية:



كيف يستجيب للإحباط أو المواقف غير السارة؟



شكل (٨ـ١٣) يوضح أمثلة لبعض موازين التقدير البيانية

ومن ميزات هذا النوع من موازين التقدير سهولة تسجيل التقديرات ، وإمكانية تقييم عدد من الخصائص أو السمات في صفحة واحدة مع جعل الميزان مشتركًا بين الخصائص جميعها، وبذلك يمكن رسم صفحة نفسية (بروفيل) لهذه الخصائص لكل فرد.

وقد أوضح جيلفورد Guilford بعض القواعد التى يمكن الاسترشاد بها فى إعداد موازين التقدير البيانية نوجزها فيما يلى:

(١) ينبغى أن تشغل كل سمة أو خاصة صفحة مستقلة ، حيث إن هذا يسمح بتقدير كل سمة على حدة قبل الانتقال إلى السمة التالية.

(٢) يجب ألا يقل طول الخط عن ١٢ سم ، ولكن لا ينبغى أن يزيد عن ذلك كثيرًا لكى يتسنى استيعابه.

(٣) يجب أن يكون الخط متصلا دون فجوات، حيث إن الفجوات تعطى انطباعًا بأن المتغير منفصل.

(٤) ينبغى تبادل الكلمتين الممثّلتين لطرفى الخط بطريقة عشوائية لتجنُّب النزعة الآلية لوضع علامة على أحد الطرفين دون الآخر.

(٥) يُفضَّل استخدام ٣ أو ٥ كلمات وصفية على الخط؛ اثنان منها طرفيتان، والباقى فى الوسط.

(٦) ينبغى أن تكون الكلمات الوصفية موجزة وحروفها صغيرة، بحيث يكون هناك فراغ بين كل كلمتين.

(٧) ينبغى استخدام كلمات وصفية متعارف عليها وأن تكون مصاغة صياغة لغوية صحيحة.

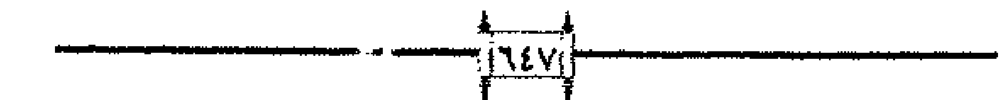
(٨) لا يجب أن تكون الكلمتان الطرفيتان متطرفتين جدًا فى معناهما لكى لا يتجنَّبهما القائم بالملاحظة.

(٩) ينبغى أن تكون الكلمات الوصفية المحايدة مقابلة لمنتصف الخط.

موازين تقدير جبرية Forced - Choice Rating Scales :

سبق أن أوضحنا فى الفصلين الحادى عشر والثانى عشر أن موازين التقدير الجبرية اقترحت لتقليل تزييف الاستجابات فى مقاييس الميول والاتجاهات.

وتُستخدم هذه الموازين لتقليل تحيُّز القائم بالملاحظة فى تقديره لخصائص وسمات شخصية الأفراد. ومثال ذلك نزعته إلى إعطاء تقديرات مرتفعة أو منخفضة لفرد معين؛ لذلك تشتمل كل فقرة من فقرات الميزان على أربع عبارات اثنين منها متكافئتان فى القبول الاجتماعى، والاثنان الأخريان متكافئتان فى عدم القبول الاجتماعى، غير أن كل عبارتين يختلفان فى درجة تمييزهما بين مجموعات محكة معينة Criterion Groups. ونظرًا لأن القائم بالملاحظة لا يكون على دراية بالعبارات



المميزة، فإنه يصعب تحيز تقديراته، حيث يُطلب منه وضع علامة على العبارة التى تصف الفرد بدرجة أفضل، والعبارة التى تصفه بدرجة أقل، وبذلك يتناول أكثر من سمة فى آن واحد. ومثال ذلك فى تقدير أداء المعلم ما يلى:

- يتحلى بالصبر مع الطلاب بطيئى التعلم.
- يثق فى نفسه أثناء الشرح.
- يهتم ويُتابع طلاب الفصل.
- يُخبر الطلاب مقدمًا بأهداف الدرس.

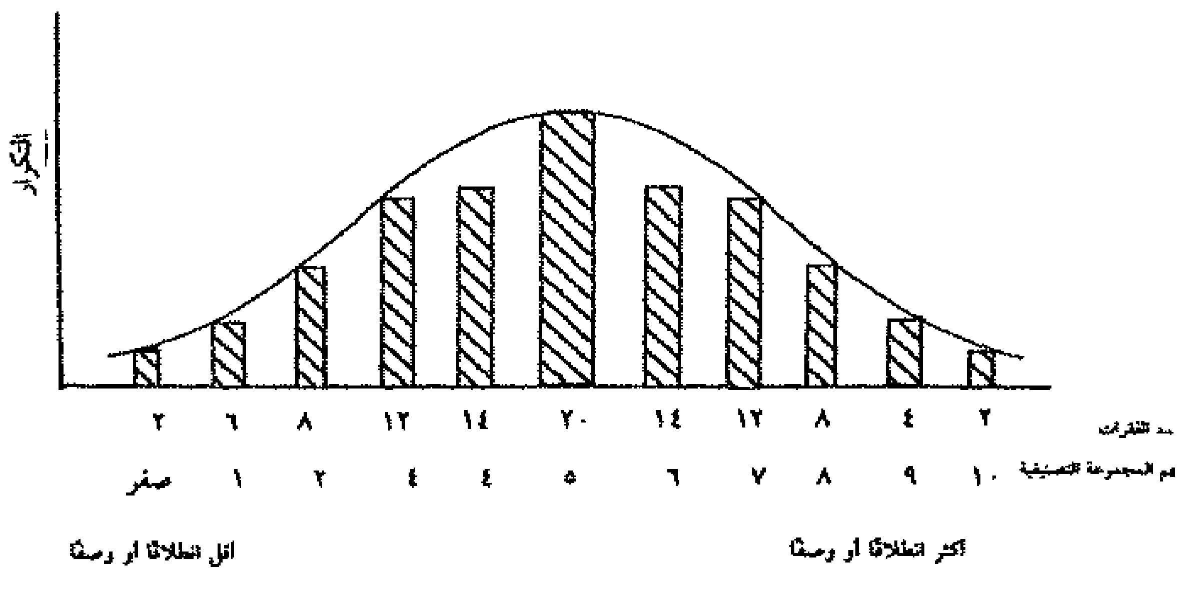
ومثال آخر:

- لا يعرف متى وأين يتدخل فى الحوار.
- يتحمل المسئولية بسهولة.
- لا يستمع لاقتراحات الآخرين.
- يُقدم مقترحات بنَّاءة.

وقد سبق أن أوضحنا بالتفصيل مزايا وعيوب هذا النوع من موازين التقدير.

وتُقدَّر الدرجات بتعيين درجة واحدة موجبة (+١) لكل عبارة مقبولة اجتماعيا وصادقة، ودرجة سالبة واحدة (-١) لكل عبارة غير مقبولة اجتماعيًا وصادقة.

ويوجد أسلوب آخر بديل لموازين التقدير الجبرية سبق أن ناقشناه فى الفصل الثانى عشر وهو أسلوب التصنيف الترتيبى Q Sort الذى اقترحه ستيفنسون (Stephenson, 1953)،ويُستخدم فى أغلب الأحيان فى تقييم مفهوم الذات Self Concept،حيث يُعطى القائم بالتقدير عددًا من العبارات الوصفية المتعلقة بالشخصية، ويُطلب منه تصنيفها فى مجموعات مرتبة ترتيبًا تنازليا وفقًا لمدى انطباقها على الفرد أو وصفها له أو شيوعها. وينبغى أن يضع عددًا معينًا من الفقرات أو العبارات فى كل قسم بحيث يكون توزيع التقديرات اعتداليا كما هو موضح بشكل (١٣-٩) التالى:



شكل (١٣ـ٩) يوضح التصنيف الترتيبى Q Sort لعدد قدره ١٠٠ فقرة أو عبارة

ويتضح من شكل (١٣ـ٩) أن معظم الفقرات أو العبارات تقع وسط التوزيع ، ويقع عدد قليل منها فى كل من طرفيه.

موازين تقدير قوائم مراجعة Checklist Rating Scales :

يُعد هذا النوع من موازين التقدير أكثر فاعلية فى تقييم الخصائص والسمات الشخصية. فالقائم بالملاحظة يضع علامات على جميع العبارات التى تشتمل عليها قائمة مراجعة والتى يتصف بها فرد معين. وقد سبق أن أوضحنا هذه القوائم فى الفصل الحادى عشر.

وعلى الرغم من أن العبارات تكون مُرتبة عشوائيًا فى القائمة، إلا أنها تتراوح بين عبارات مفضلة وعبارات غير مفضلة وتتعلق بسمات متعددة. وللحصول على الدرجة الكلية للتقدير، يوجد متوسط درجات تقديرات جميع العبارات. ويمكن استخدام أسلوب الفترات المتساوية ظاهريًا الذى اقترحه ثيرستون فى تحديد القيم المناظرة لكل عبارة. وقد ناقشنا ذلك بالتفصيل فى مجال قياس الاتجاهات فى الفصل الثانى عشر.

كما يمكن استخدام هذه القوائم فى تقدير المشكلات الشخصية وذلك فى مجال الإرشاد المدرسى من أجل التعرف على الطلاب الذين لديهم مشكلات شخصية معينة. ومن أمثلة ذلك قوائم مونى للمشكلات Mooney Problem Checklists ، وتتكون هذه القوائم من أربع صيغ للمستويات العمرية المختلفة، ويشتمل كل منها على عدد من

المشكلات يتراوح بين ٢١٠ ، ٣٣٠ مشكلة فى المجالات الصحية، والجسمية، والنمو، والمنزل، والأسرة، والعلاقات بين الجنسين، والأخلاقيات، والتدين، والزواج، والأمن الاقتصادى، والمدرسة أو العمل، والمجالات الاجتماعية والترويحية. ويطلب من الفرد وضع خط تحت المشكلات الأقل أهمية، ويضع دائرة حول المشكلات الأكثر أهمية، ويكتب ملخصًا عن مشكلاته. وتقدر درجات تناظر عدد المشكلات فى كل من هذه المجالات، وتفسر الاستجابات بطريقة ذاتية، حيث إنه لا توجد معايير لهذه القوائم.

ومن الأمثلة الأخرى قائمة الصفات The Adjective Checklist (ACL) التى أعدها جوخ ، وهيلبرن (Gough & Heilburn , 1965)، وتشتمل على ٣٠٠ صفة مرتبة أبجديًا، وتتعلق بمشكلات شخصية، مثل شارد الفكر، ومضحك،ويستغرق تطبيقها ٢٠ دقيقة. وتختلف عن قوائم مونى بأن تفسير درجاتها يستند إلى معايير الدرجات التائية، وتعد هذه القوائم من نوع التقرير الذاتى.

وكذلك قائمة سلوك الطفل Child Behavior Checklist (CBCL) التى أعدها أشنباك، وإدلبروك (Achenbach & Edelbrok, 1991) لكى يستخدمها الآباء فى تقدير الكفايات والمشكلات السلوكية لأطفالهم، حيث قاما بتحديدها باستخدام التحليل العاملى.

وفيما يلى أمثلة أخرى لبعض موازين تقدير السلوك المتوافرة فى مجال تقييم الشخصية فى الصف المدرسى:

(١) بروفيل تقدير السلوك Behavior Rating Profile (BRP-2) :

أعده براون، وهاميل (Brown & Hammill , 1990) واستندا فى بنائه على المدخل البيئى Ecological Approach، حيث حصلا على معلومات من مصادر بيئية متعددة. ويشتمل على موازين تقدير للطلاب والمعلمين والآباء والعلاقات الاجتماعية، وبذلك يمكن تقويم الطالب بطريقة فردية فى مواقف متعددة بواسطة عدد من المحكمين، ويناسب الأعمار التى تتراوح بين ٦,٥ ، ٨,٥ عامًا.

(٢) نظام تقدير المهارات الاجتماعية

: Social Skills Rating System (SSRS)

أعده جريشام، وإليوت (Gresham & Elliot , 1990) ويعد بمثابة بطارية تسمح بتقديرات متعددة للسلوك الاجتماعى الذى يمكن أن يؤثر فى علاقات المعلم بالطلاب، وعلاقات الطلاب فيما بينهم، والتحصيل الدراسى. وتتكون البطارية من صيغتين؛

إحداهما تتعلق بالمعلمين والآباء، والأخرى تتعلق بالطلاب في مختلف المراحل التعليم العام.

(٣) ميزان تقدير الكفاية الاجتماعية والتكيف المدرسي لوالكر ومكونل

Walker - McConnell Scale for Social Competence and School Adjustment (W-M Scale):

وقد أعدّاه عام ١٩٨٨ لتعرُّف الأطفال الذين يفتقرون إلى المهارات الاجتماعية، ويستغرق تطبيقه خمس دقائق فقط. ويشتمل على ثلاثة موازين تقدير فرعية تتعلق بالسلوك الاجتماعي الذي يُفضّله الأطفال لدى معلميهم، ولدى أقرانهم، والتكيف المدرسي. ويمكن أن يُطبقه المعلم الذي أتيحت له فرصة ملاحظة أطفال صف معين مدة شهرين على الأقل، وتوجد معايير لهذه الموازين من الصف الأول إلى السادس.

(٤) موازين تقدير السلوك لبيركس

Burks Behavior Rating Scales (BBRS):

أعدّها بيركس (Burks, 1977) لتعرُّف بعض أنماط السلوك غير المقبول للأطفال في المرحلتين الابتدائية والإعدادية، وذلك بهدف الحصول على معلومات تفيد المعلم وأولياء الأمور في تعرف السلوك المضطرب لدى بعض الأطفال، والتنبؤ بأدائهم في فصول الفئات الخاصة

وتُستخدم موازين تقدير قوائم المراجعة إذا كنا نهتم بمعرفة ما إذا كانت خاصة أو سمة معينة متوفرة أم لا. لذلك ينبغي تحديد الخصائص أو السمات المرجو ملاحظتها، وأن يكون لكل فرد قائمة مراجعة خاصة به، ويتم ملاحظته فرديًا في ضوء العبارات أو الفقرات التي تشتمل عليها القائمة.

وقد سبق أن أوضحنا في الفصل السابق بعض الإرشادات العامة لبناء موازين التقدير.

مصادر الخطأ في موازين التقدير، وكيفية التغلب عليها:

هناك العديد من مصادر الخطأ الشائعة في أساليب التقدير، بعضها يرجع إلى موازين التقدير، وبعضها الآخر يرجع إلى القائم بعملية التقدير وإلى السمات المراد تقديرها. وفيما يلي توضيح لأهم مصادر هذه الأخطاء:

(١) تأثير الهالة Halo Effect : ويحدث هذا الخطأ عندما يؤثر الانطباع العام للقائم بالتقدير عن الفرد في كيفية تقديره له في سمات مختلفة. فإذا كان لديه انطباع جيد بعامة عن فرد معين، فإنه يميل إلى إعطائه تقدير مرتفع في معظم السمات،

وعكس ذلك صحيح . ويؤدى هذا الخطأ إلى قيم مرتفعة زائفة لمعامل الارتباط بين سمات تبدو غير مرتبطة . كما تؤدى إلى إخفاء الفروق الفردية ، وذلك لأن تأثير الهالة يجعل التقديرات متراكمة عند طرفى ميزان التقدير ، مما يقلل من صدق التقديرات .

ويمكن التغلب إلى حد ما على هذا المصدر من الخطأ بتعريف السمات تعريفا إجرائيًا دقيقًا، وكذلك بتبادل الكلمتين أو الصفتين الممثلتين لطرفى الميزان بطريقة عشوائية فى موازين التقدير البيانية كما سبق أن أشرنا .

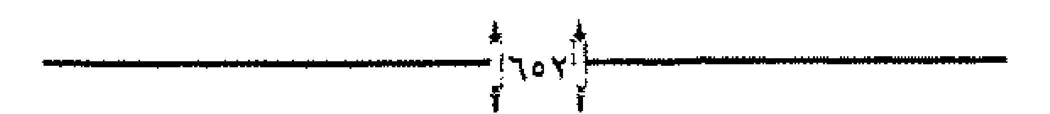
(٢) التساهل فى التقديرات Leniency : ينزع بعض القائمين بعملية التقدير إلى التساهل فى تقديراتهم للأفراد ، وبذلك يتجنّبون استخدام الطرف الأيسر الذى يُمثل القيمة المنخفضة على ميزان التقدير . فكثير منهم لا يفضّل الحكم على سمة معينة لدى الفرد بأنها ضعيفة جدًا لكى لا يبدو متشددًا . ويمكن أيضًا أن يحدث العكس حيث ينزع البعض إلى التشدد فى التقديرات ، وبذلك يتجنب استخدام الطرف الأيمن لميزان التقدير .

فالخطأ فى كلتا الحالتين يُعد خطأ متسقًا فى اتجاه تحيز القائمين بعمليتى التقدير . وهذا المصدر من مصادر الخطأ لا يكون تأثيره ذا أهمية إذا قام بعملية تقدير السمات نفسها فرد واحد أو عدة أفراد ، حيث إن اتجاه تحيزهم يكون معلومًا ، غير أنه إذا قام عدة أفراد بعملية التقدير ، وأُجريت مقارنة بين تقديراتهم ، فإنه ينبغى تعرف اتجاه تحيزهم لكى يؤخذ بعين الاعتبار .

وهذا المصدر من مصادر التحيز يقلل من مدى ميزان التقدير مما يقلل بالتالى من تمييز التقديرات . ويمكن التغلب على ذلك باستخدام موازين التقدير جبرية الاختيار ، أو جعل عدد الأقسام الأكثر تفضيلا أقل من عدد الأقسام الأقل تفضيلا على ميزان التقدير ، وفى بعض الأحيان يمكن استخدام موازين التقدير البيانية بدلاً من الموازين جبرية الاختيار .

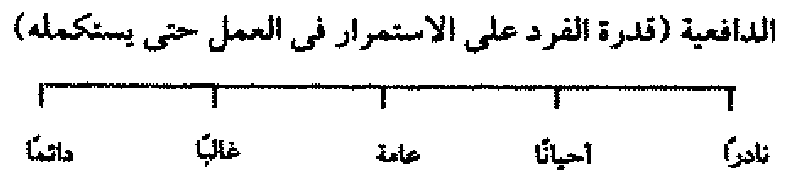
(٣) النزعة المركزية Central Tendency : ويحدث هذا الخطأ عندما يتجنب القائم بالتقدير طرفى الميزان ويُركز تقديراته فى الوسط ، وهذا يقلل من الفروق الفردية فى تقديرات الأفراد مما يقلل بدوره من ثبات هذه التقديرات . ويمكن التغلب على ذلك أيضًا باستخدام الموازين جبرية الاختيار ، وكذلك بتوضيح الغرض من عملية التقدير للقائمين بها بحيث يعملون على إبراز الفروق الفردية للتقديرات بدلا من العمل على تلافيها .

(٤) غموض السمات المراد تقديرها Ambiguity Error : ويحدث هذا الخطأ عندما تكون السمات المراد تقديرها غير معرّفة تعريفًا جيدًا ، فسمات مثل الدافعية ، وسوء التكيف ، والانقباض ربما تُفسّر تفسيرات مختلفة . كما أن الكلمات الوصفية التى



تحدد أقسام موازين التقدير البيانية ربما لا تكون واضحة المعنى، فالفرق بين «بدرجة كبيرة» ، و«مرتفع» غير واضح. وللتغلب على ذلك ينبغى العناية بتعريف السمات والخصائص تعريفًا دقيقًا بحيث يستند القائمون بالتقدير إلى إطار مرجعى محدد. كما يمكن وضع التعريف إلى جوار السمة على ميزان التقدير، وانتقاء الكلمات الوصفية بحيث لا تحتمل أكثر من معنى واحد كما يلى:

الدافعية (قدرة الفرد على الاستمرار فى العمل حتى يستكمله)



(٥) عدم تدرب القائم بالتقدير، ترجع مصادر الخطأ السابقة فى معظمها إلى عدم استخدام القائمين بالتقدير محكات منظمة. وكذلك إلى عدم تدربهم على مهارات الملاحظة والتقدير وتسجيل البيانات، وهذا يؤدى إلى انخفاض ثبات التقديرات. ويمكن التغلب على ذلك بتنظيم دورات تدريبية للقائمين بالملاحظة وذلك لتوضيح أهمية تحرِّى الدقة والأمانة فى التقدير، والتعريف بمصادر الخطأ الشائعة وكيفية تجنبها أو التقليل منها، وتطبيق ذلك فى بعض المواقف المماثلة للواقع الفعلى.

كما ينبغى توضيح معنى السمات أو الخصائص المراد تقديرها فى عبارات وصفية سلوكية يمكن ملاحظتها بحيث لا تتطلب إلا قدرًا ضئيلاً من الاستدلال. ولعل أسلوب الأحداث الحرجة Critical Incidents Technique الذى أوضحناه فى الفصل العاشر يساعد فى تحديد السلوك الأكثر أهمية الذى ينبغى ملاحظته. وكذلك توضيح أقسام ميزان التدريج ودلالتها، والتدرب على استخدامها فى الملاحظة من خلال ورش عمل منظمة.

ابما أساليب المقابلة Interview Techniques :

تُعد المقابلة من أكثر الأساليب المستخدمة فى تقيم الشخصية، وتُزودنا بمعلومات تختلف عما تزودنا به أساليب ملاحظة السلوك، حيث إنه فى المقابلة يتحدث الفرد عن نفسه، بينما فى الملاحظة يتم تقدير سلوكه. فالمقابلة تُستخدم فى أغراض متعددة فى مجالات الإرشاد، والمجالات الكلينيكية، والتربوية، والاجتماعية، والصناعية، والقانونية، واستطلاع الآراء، وكذلك فى اتخاذ قرارات تتعلق بانتقاء الأفراد وترقيتهم، وفى تقويم البرامج التربوية لجمع بيانات تتعلق بخلفية الطلاب وتعرُّف وجهة نظرهم حول برنامج معين، واستطلاع آراء المعلمين والإداريين حول محتوى وأسلوب تنفيذ البرامج التعليمية، كما تُستخدم فى الأغراض البحثية.

والمقـابلة تتضمن تفـاعلاً لفظيًّا شـفويًا وجهًا لـوجه بين فردين أو أكثـر ، حيث يحاول أحدهم وهو القائم بالمقابلة Interviewer أن يتعرف مثلا على مفهوم الذات لدى الفـرد ، وعقائـده، ومقاصـده، وأسلوب تفـكيره، وآرائه وغـير ذلك مـن خصائـصه الشخصية.

وعلى الرغم من أن المقـابلة لا تُعد أداة قياس بالمفهوم السـيكومترى، لأنها لا تؤدى عادة إلى بيـانات كمية، إلا أنها تقدم بـيانات ثرية إذا قام بها أخصـائى نفسى أو تربوى مدرب تدربًا جيدًا، بل وأحيانًا يمكن أن تُزودنا بـبيانات كمية كما هو الحال فى أساليب الملاحظة، إذ يمكن اعتبار المقابلة اسـتبيانًا شفويًا إذا روعى تقنين الأسئلة التى تُوجَّه للفرد.

فالمقـابلة تمدنا بمعلومات كـلينيكية وفينومونـولوجية، وتعتمد هذه الـمعلومات على الإطار النظرى الذى يستند إليه القائم بالمقابـلة، فهناك أطر نظرية مختلفة للمقابلة تنسب إلى فرويد Freud ،وروجرز Rogers ،وساليفان Sullivan،وغيرهـم. فالمقابلة عند فرويد Freud التى تعتـمد على التحليل النفسى يكون الاهتـمام فيها مـركزًا على تحديد طبيعة صـراعات الفرد والكشف عن اللاشعور. وهنا يهتم الأخـصائى بالمعانى المتعددة الشعوريـة واللاشعورية المقصودة وغير المقصـودة فى مفردات لغة الفرد الذى تُجرى معه المقابلة، ويساعد الأخصائى الفرد على استرجاع ذكريات ماضيه ليساعده فى إدراك منشـأ الصراعات الـتى ربما يعـانى منها. وهـنا يعتـمد الأخصـائى فى تفسـيره للمعلومات التى جمـعها عن الفرد على أكثر مما تسمح به هذه المـعلومات، أى يعتمد على فروضـه الخاصة عن طبيـعة القوى النفسـية التى تحكم السلـوك الإنسانى، وهذه الفروض النظرية تُوجِّه عـملية المقابلة ويتم التفسـير فى ضوئها (Freud, 1953) . أما المقابلة عند روجرز Rogers التى تعتمـد على المدخل الفينومونولوجى فـتتركز حول العميل، ويسأل فيها الأخصائى عددًا قليلاً مـن الأسئلة المباشرة، ويُعيد صياغة إجابات الفرد عنهـا، ويُحثه على الاسـتمرار فى الإجابة، ويتـجنب أى تفسـيـرات لا تقوم على أساس المعلومات التى حصل عليها، أى أن الأخصـائى هنا يهتم بالطريقة التى يشعر أو ينظر بها الفرد إلى نفسه.

ويؤكد هذا المدخـل مفهوم الذات Self Concept ، وهو الجـزء من المـجال الظاهرى الـذى يتكون من تركيـبة من الإدراكات والـقيم المتـعلقة «بالذات» أو بـالفرد كمصـدر للخبرة والسـلوك، أى أن «الذات» هى المحور الرئـيسى للخبـرة. لذلك فإن

المقابلة تُعد أحد الأساليب المهمة التي يمكن باستخدامها الحصول على معلومات تتعلق بمفهوم الذات (Combs, 1959).

أنواع أساليب المقابلة:

تتنوع أساليب المقابلة وفقًا للغرض من استخدامها، غير أنه يمكن تصنيفها في نوعين رئيسين هما: المقابلة غير المقننة Unstandardized Interview، والمقابلة المقننة Standardized Interview، وتتباين المقابلة في مدى تحديد بنيتها، فهي تتراوح بين البنية المحددة تحديدًا جيدًا، والبنية غير المحددة إطلاقًا. فالمقابلة محددة البنية Structured Interview تعتمد على قائمة من الأسئلة مُعدة مسبقًا، أما المقابلة غير محددة البنية Unstructured Interview فتعتمد على مجموعة من الأسئلة، ولكن يمكن تعديلها بالحذف أو بالإضافة أثناء المقابلة. وتوجد أنواع متعددة من المقابلة بغض النظر عن مدى تحديد بنيتها، وتختلف فيما بينها من حيث الغرض الذي تُستخدم من أجله، فهناك المقابلة الشخصية لانتقاء الأفراد Personnel Interview، أو تعرُّف مشاكلهم أثناء العمل Trouble Shooting Interview، والمقابلة المتعلقة بالاقتراع Poll Interview، والمقابلة الكلينيكية Clinical Interview التي تُستخدم في المجالات الطبية والنفسية والاجتماعية، مثل المقابلة الطبية التشخيصية Medical Diagnostic Interview، ومقابلة الانعصاب أو مقابلة الاستجابة الدفاعية Defensive- Response (Stress) Inter-view، والمقابلة الطبية النفسية المدخلية Psychiatric Intake Interview، والمقابلة لتحديد الحالة العقلية Mental Status Interview، والمقابلة لدراسة تاريخ الحالة Case History Interview، وهناك أيضًا المقابلة البحثية Research Interview، مثل المقابلة المسحية لتحديد الحالة النفسية الراهنة لمجتمع من الأفراد، والمقابلة التي تتم بعد إجراء تجريبي معين Postexperiment Interview وكذلك توجد أساليب محاكاة المقابلة Interview Simulatio باستخدام برامج حاسوب يتم تصميمها لإجراء نوع معين من المقابلة الاصطناعية، وغير ذلك.

ونظرًا لضيق المجال، فإننا لن نستطيع مناقشة جميع هذه الأنواع، وإنما سوف نقتصر فيما يلي على توضيح النوعين الرئيسين السابق ذكرهما.

المقابلة غير المقننة: يُعد هذا النوع من المقابلة غير محدد البنية، حيث إن الأسئلة التي توجَّه للفرد لا تكون محددة سلفًا، وإنما تتحدد بناء على استجاباته، فهذه المقابلة تماثل الحديث العادي بين فردين، وتُستخدم عادة في المقابلات الشخصية

للمتقدمين للوظائف أو للالتحاق بالكليات. وعلى الرغم من أنه يصعب الحصول على بيانات كمية استنادًا إلى المقابلة غير المقننة، إلا أنها تتميز بتعدد الاستجابات نظرا لمرونة الأسئلة، وتسمح للقائم بالمقابلة أن يختار المجالات التى يهتم بها الفرد الذى تُجرى معه المقابلة والتى لم تكن متوقعة، أى أنها تهتم بفردية الشخص وتجعله فى حالة ارتياح ولا يلجأ إلى تمثيل الأدوار، وبذلك تساعد فى التحليل الفينومونولوجى. غير أنه يُعاب عليها صعوبة المقارنة الكمية بين الأفراد أو بين الأسئلة التى تُقدَّم لهم نظرًا لأنها تختلف باختلاف المواقف والقائمين بالمقابلة، ولكن يمكن معالجة البيانات معالجة كيفية أو تحليلها تحليلا موضوعيًا. كما يمكن أن يعتمد التفسير اعتمادًا مباشرا على ما تم التوصل إليه من معلومات، أو يمكن أن يتضمن التفسير فروضًا أبعد مما تسمح به هذه المعلومات.

وهذه المعالجة الكيفية لها أهمية فى عملية تحليل المحتوى Content Analysis التى تُستخدم فى تصنيف هذه المعلومات بأسلوب منظم، وتختلف طبيعة هذا التصنيف باختلاف محتوى المقابلة والغرض منها.

المقابلة المقننة: على الرغم من أن هذا النوع من المقابلة يُستخدم فيه نمط الأسئلة نفسها كما فى المقابلة غير المقننة، ويهدف لجمع معلومات مماثلة من كل فرد؛ إلا أن الأسئلة التى تُوجَّه لكل منهم تكون مُحددة قبل بدء المقابلة، وتكون موحَّدة فى صياغتها ، وتُقدم بالترتيب نفسه لجميع الأفراد، لذلك تُعرف باسم «المقابلة وفق جدول Schedule Interview» . وبذلك تيسِّر المقابلة المقننة عملية القياس والتقدير الكمى للاستجابات، وتسمح بالمقارنة بين الأفراد، والبيانات المستمدة منها ربما تكون أكثر ثباتًا من النوع الأول.

وأحيانًا تُستخدم موازين التقدير Rating Scales التى أوضحناها منذ قليل إلى جانب المقابلة المقننة كوسيلة لتلخيص المعلومات التى يتم الحصول عليها، وعادة يسجَّل ما دار من حوار أثناء المقابلة، ومن هذا يتضح أن نقاط قوة المقابلة المقننة تُعد نقاط ضعف للمقابلة غير المقننة، والعكس صحيح.

مزايا وعيوب أساليب المقابلة:

تتميز المقابلة بنوعيها بأنها تتيح الفرصة للفرد أن يُعبرَّ عن نفسه ويتساءل، كما تتميز نسبيًا بسهولة الاستفسار من الأطفال والأميين، والحكم على أسلوب تعبيرهم إلى جانب محتوى المقابلة. غير أنه يُعاب عليها صعوبة تحليل وتصنيف الاستجابات غير محددة البنية، وكلفة تدريب القائمين بالمقابلة، وإجراء المقابلة مع كل فرد على حدة.

ومهما بلغت فاعلية التدريب فإن الخصائص الشخصية للقائم بالمقابلة مثل معتقداته وتوقعاته ونزعاته ربما تؤثر تأثيراً منظماً في استجابات الأفراد، وهذا يُعد مصدراً من مصادر الخطأ في المقابلة، ويُطلق عليه الخطأ المتنقل Contagious Bias ، كما أنه ربما يسيء فهم ما يقوله الفرد أثناء المقابلة، أو أن الفرد يسيئ فهم ما يقوله القائم بالمقابلة. غير أنه يُمكن تقليل أثر تحيز القائمين بالمقابلة بالعناية بتصميم إجراءات المقابلة ومحتواها، وكذلك بتدريبهم لتنمية مهارات المقابلة لديهم، وجعلهم على دراية بالمشكلات المحتملة، وأن يتم انتقاؤهم بحيث يكون هناك قدر من الاتفاق بين خصائصهم وخصائص الأفراد الذين سوف تجرى المقابلة معهم. فكلما قلت مصادر الخطأ والتحيز في عملية المقابلة ، كانت البيانات المستمدة منها أكثر صدقاً.

تقييم أساليب الملاحظة والمقابلة :

يتضح مما سبق أن أساليب الملاحظة والمقابلة تهدف للحصول على معلومات مهمة تتعلق بخصائص أو سمات شخصية الفرد، على الرغم من أن أساليب الملاحظة تعتمد على ما يفعله الفرد، بينما تعتمد أساليب المقابلة على ما يقوله الفرد عن نفسه؛ لذلك فإن كلا من هذه الأساليب شائع الاستخدام في التقييم المنظم لخصائص الأفراد. ويمكن اعتبار عمليات الملاحظة والمقابلة فناً وعلماً، فبعض القائمين بها يُجرون هذه العمليات بفاعلية أكثر من غيرهم، ويكون لديهم مهارة في تكوين علاقة وئام بينهم وبين الأفراد الذين يلاحظونهم أو يجرون مقابلة معهم. وعلى الرغم مما تتميز به هذه الأساليب من مرونة وواقعية وما تؤدي إليه من بيانات واستجابات ثرية ومتنوعة ربما يصعب الحصول عليها باستخدام الاستبيانات أو الأساليب الإسقاطية، إلا أنها أيضاً لا تخلو من مشكلات متعددة ، وهذه الأخطاء إذا لم نكن مدركين لها ونعمل على تجنبها أو تقليل أثرها، فإنها تؤدي إلى خفض ثبات وصدق البيانات أو المعلومات التي نحصل عليها.

وهذا لا يتأتى إلا إذا اعتُني بتصميم أساليب الملاحظة وما يتعلق بها من موازين تقدير، وكذلك أساليب المقابلة ، مع مراعاة تقنين هذه الأساليب، والاستناد إلى متوسط تقديرات عدد ممن يقومون بالملاحظة لكي تتلاشى عوامل التحيز، وبذلك نحصل على صورة أكثر دقة لخصائص شخصية الفرد، وبالتالي يزداد ثبات التقديرات. فزيادة عدد القائمين بتقدير فرد أو عدة أفراد يماثل زيادة عدد مفردات الاختبار مما يؤدي إلى رفع قيمة الثبات، فالثبات يشير إلى استقرار أو اتساق التقديرات ، أو اتفاق القائمين بالمقابلة.

وتتباين قيم الثبات بتباين القائمين بالملاحظة أو المقابلة ومدى اتفاقهم، إذ وجد بعض الباحثين أن هذه القيم تتراوح بين ٠,٢٣ ، ٠,٩٧ ، ووسيط هذه القيم ٠,٥٧ (Ulrich & Trumbo , 1965). وربما يرجع هذا التباين إلى أن القائمين بالملاحظة أو المقابلة ينظرون إلى أشياء مختلفة، إذ ربما ينظر أحدهم إلى جوانب قوة الفرد، بينما ينظر آخر إلى جوانب ضعفه. فاختلاف تقديراتهما في هذه الحالة يرجع إلى أن كلا منهما تستند أحكامه إلى جوانب مختلفة للفرد. كما أن التوجه النظرى لكل منهما يؤثر فى نوع المعلومات التى يبحث عنها والتى يستند إليها فى اتخاذ القرار.

أما مصادر الخطأ المتسقة مثل تأثير الهالة، وتأثير هاوثورن Hawthorn Effect الذى يؤدى إلى تغيير استجابات الأفراد، وغير ذلك من الأخطاء فإنها تؤدى إلى خفض صدق البيانات المستمدة من الملاحظة والمقابلة،حيث إن الصدق يتعلق بمعنى الدرجة. فالأخطاء التى تقلل من موضوعية الملاحظة أو المقابلة تؤدى إلى أحكام غير دقيقة، وبالتالى فإنها تؤدى إلى تحيز فى معنى التقديرات أو صدق التقييم، وربما يفسر ذلك تباين قيم الصدق التنبؤى لهذه الأساليب.فقد أوضح ألريك Ulrich وترامبو Trumbo أن قيم معامل الصدق التنبؤى فى مجموعة من الدراسات تراوحت بين - ٠,٠٥ و ٠,٧٢ باستخدام محكات مختلفة ، مثل النجاح فى الكليات، والأداء فى العمل، وأكد ذلك دراسة كارلسون Carlson وزملائه (Carlson, et al. , 1971) . وعمومًا فإن صدق هذه الأساليب يختلف باختلاف الغرض من استخدامها، وطريقة تسجيل المعلومات المستمدة منها. فكلما كانت هذه الأساليب غير مقننة قل ثبات نتائجها وصدقها، ومع هذا فإنها تزودنا عادة بمعلومات ربما يصعب الحصول عليها من مصادر أخرى، ولكن الاعتماد على هذه المعلومات غير المتسقة أو منخفضة الصدق سوف يكون محدودًا؛ لذلك يفضَّل أن تكون أساليب الملاحظة والمقابلة مقننة، وتُستخدم فى أيد ماهرة إذا كان اتساق النتائج يُعد أمرًا ضروريًا.

ومن الجدير بالذكر، أنه لا ينبغى استخدام أساليب المقابلة فى التحقق من صحة فروض بحثية معينة، وإنما يمكن استخدامها فى الملاحظات الكشفية لاشتقاق فروض يمكن بحثها بأساليب أفضل (Guilford, 1971) . وعلى الرغم من هذا كله، فإن أساليب الملاحظة والمقابلة سوف تظل من الأدوات الأساسية لتقييم الشخصية. وبالطبع تزداد فائدتها فى المواقف الكلينيكية، ومواقف الإرشاد، وفى انتقاء الأفراد، ومواقف العمل حيث يكون لها استخدامات مختلفة.

(رابعاً) مقاييس محددة البنية وغامضة الهدف

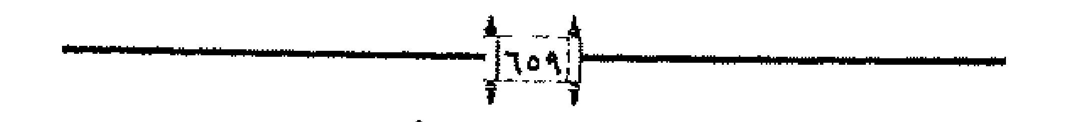
(مقاييس الأنشطة الفسيولوجية، ومقاييس الإدراك) :

اتضح لنا مما سبق أن استبيانات الشخصية، والأساليب الإسقاطية، وأساليب الملاحظة والتقدير، والمقابلة، تُعد أدوات أساسية لتقييم الشخصية، غير أن هناك أساليب أخرى تعرف بالأساليب الموضوعية Objective Techniques، وتستند إلى توجهات مختلفة لا تزال موضع دراسة وبحث. فأساليب التقرير الذاتي، والأساليب الإسقاطية تؤكد جوانب الاختلاف في السلوك الإنساني، وتُركز على الوظائف العامة في شخصية الفرد، غير أن الأدوات أو المقاييس التي ستتناولها بإيجاز فيما يلي تركز على الأنشطة التي يمكن ملاحظتها، وعلى الموقف المحيط بها، فهذه المقاييس لا تعتمد على تقديم سؤال للفرد بطريقة مباشرة مما يتطلب تقريراً ذاتيًا كما في الاستبيانات، ولا تعتمد على الافتراض الذي تستند إليه الأساليب الإسقاطية ، أو المثيرات غير محددة البنية. فهذه المقاييس غير مباشرة، وواضحة الهدف وموضوعية. وأمثلة ذلك مقاييس الأنشطة النفسية الفسيولوجية، والمقاييس الموضوعية للأداء، ومقاييس القدرة الإدراكية التي تكون محددة البنية.

وتُعد كثير من هذه المقاييس اختبارات أداء، حيث يوضع الأفراد في موقف مقنن، أي يُعرض المثير نفسه على جميع الأفراد في الظروف نفسها في المختبَر، وتُستخدم الملاحظة المباشرة في الحصول على معلومات دقيقة يمكن الاستناد إليها. ولكي يتم ضبط الموقف تكون هذه الاختبارات محدودة في نوع وعدد المثيرات التي تقدَّم للفرد، وفي خصائص الشخصية التي يُفترض قياسها.

فالهدف من هذه الاختبارات الحصول على معلومات تتعلق بسلوك معين في ظروف مقننة. وتهتم النظريات التي تؤكد استخدام هذه الأنواع من الاختبارات الموضوعية والمحددة البنية بالعمليات المشتركة بين جميع الأفراد . ويرى المؤيدون لها أن علم الشخصية يمكن أن يتقدم بدرجة أفضل بالدراسة المنظمة لخصائص الأداء في مواقف محددة ومعرَّفة تعريفًا جيدًا . وهذا لا يقلل بالضرورة من أهمية فهم العلاقات بين الأجزاء عند تقييم الشخصية، ولكن البحث المنظم لهذه الأجزاء يجب أن يسبق أي محاولة لفهم الشخصية (Goldfried & Kent, 1972).

وتعتمد هذه النظريات إلى حد كبير على البحوث النفسية العامة في مجالات علم النفس الاجتماعي، وعلم النفس المعرفي، وبخاصة الإدراك والتعلم، وأحيانًا يصعب



التمييز بينها كنظريات فى الشخصية وبين النظريات العامة للسلوك. فقد اهتمت هذه النظريات بالخصائص العامة للأفراد، كما اهتمت بأجزاء السلوك بدلاً من اهتمامها بالسلوك ككل.

مقاييس الأنشطة الفسيولوجية :

تُستخدم مقاييس الأنشطة أو الحالة الفسيولوجية Physiological Measures فى تقييم التغيرات التى تحدث فى الاستجابات نتيجة للإثارة أو لمواقف ضاغطة. فأنماط الموجات الكهربائية للمخ، ونبض القلب، وضغط الدم، والاستجابة السيكوجلفانية للجلد، ونشاط الغدة الدرقية، وقدرة العضلات ، ودرجة حرارة الجسم، وكيمياء الدم، ترتبط بالإثارة الانفعالية واليقظة أو تركيز الانتباه ، ويختلف الأفراد فى النمط المميز لهذه الأنشطة الفسيولوجية. فجهاز كشف الكذب Polygraph يُستخدم فى تسجيل بعض هذه الأنشطة الفسيولوجية فى آن واحد فى موقف مقنن أثناء المقابلة أو أثناء قراءة كلمات أو عرض أحداث معينة مرتبطة بجريمة اتهم فيها الفرد، أو سؤاله، ويكشف فقط الاستجابة الانفعالية التى تنجم عن ذلك. وإذا استخدم الجهاز أخصائى ماهر، فإنه يمكنه تعرُّف ما إذا كان الفرد بريئًا أو مذنبًا. غير أن القضاء الأمريكى يرى أن هذا الأسلوب ليس صادقًا بدرجة تسمح بتقديم نتائجه كأدلة للحكم.

وجهاز رسَّام المخ الكهربائى Electroencephalogram (EEG) الذى يُستخدم فى تخطيط موجات المخ الكهربائية يُعد أيضًا من أدوات التقييم النفسى الفسيولوجى فى المجالات الكلينيكية للكشف عن إصابات المخ، وكذلك فى طب الأعصاب.

ويرى كثير من الباحثين أن الفروق الفردية فى مستوى الوظائف الفسيولوجية، أو فى مقدار التغير الذى يحدث فيه يرجع إلى عوامل تشريحية وليس إلى عوامل نفسية. لذلك يُفضِّلون استخدام المؤشرات الفسيولوجية فى المقارنة بين استجابات الأفراد لمثيرات متنوعة لمعرفة ما يسيطر على انتباه الفرد ، وما يستثيره، وما إلى ذلك (Averill & Opton , 1968). غير أنه قد بينت الدراسات على الشخصيات الفصامية أن استجاباتهم الفسيولوجية للمواقف الضاغطة Stress أكبر من استجابات الأسوياء.

ومع هذا، فإن نتائج البحوث بعامة أوضحت عدم وجود أنماط فسيولوجية متمايزة تقترن بانفعالات معينة، أو بخصائص مختلفة للشخصية.

مقاييس الإدراك :

تُعد هذه المقاييس حلقة وصل بين مقاييس القدرات ومقاييس الشخصية، وقد أمكن التوصل إليها من تجارب علم النفس المعرفي في مجال المعرفة Cognition، والإدراك Perception. فقد لاحظ علماء النفس أن هناك فروقًا فردية اتضحت لدى المشاركين في التجارب المتعلقة بتجهيز المعلومات Information Processing، والانتباه Attention.

وأطلق هؤلاء العلماء تسميات مختلفة على هذه الفروق ، غير أنهم أشاروا جميعًا إلى الفروق الشخصية في استراتيجيات الضبط، وأطلقوا عليها مصطلح «الأساليب المعرفية Cognitive Styles» الذى يشير إلى كل من القدرة والدافعية. وقد أوضحت دراسات تجريبية متعددة أن هناك علاقات دالة بين الخصائص المتعلقة باتجاهات الفرد ودوافعه وانفعالاته، وبين أدائه في مهام إدراكية أو معرفية، ويُعد اختبار رورشاخ الذى سبق أن أوضحناه واحدًا من الاختبارات الإدراكية Perceptual Test.

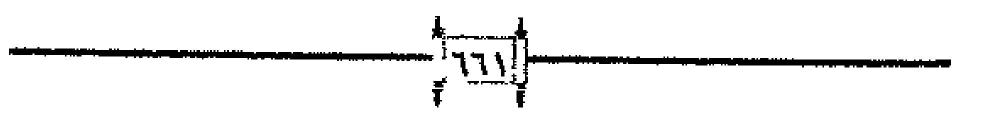
وتعد دراسات هيرمان ويتكن ومعاونيه (Witkin, et al., 1954,1967,1975) من الجهود البارزة في هذا المجال. فقد استطاعوا تحديد أحد الأبعاد الرئيسة للأساليب المعرفية أطلقوا عليه مصطلح «الاعتماد ـ الاستقلال عن المجال» Field Dependence VS. Field Independence، ولكنهم أطلقوا عليه بعد ذلك مصطلح التمايز السيكولوجى Psychological Differentiation.

ويشير كلا المصطلحين إلى الفروق بين الأفراد فى أساليب انتقاء وتنظيم موقف أو مجال معقد لإنتاج استجابة معرفية تُطلب منهم. فالفرد المستقل عن المجال أو المتمايز يمكنه فصل الأجزاء المطلوبة من المثير المربك الذي يُقدم له، أما الفرد المعتمد على المجال أو غير المتمايز فانه سيستجيب للموقف ككل دون تحليله، مما يدل على اهتمامه بالموقف ذاته. وقد توصل ويتكن Witkin إلى هذا البُعد المعرفى عن طريق مجموعة مبتكرة من الاختبارات التى يبدو أنها مترابطة وتُستخدم فى تقييم هذا الأسلوب المعرفى. وهذه الاختبارات هى:

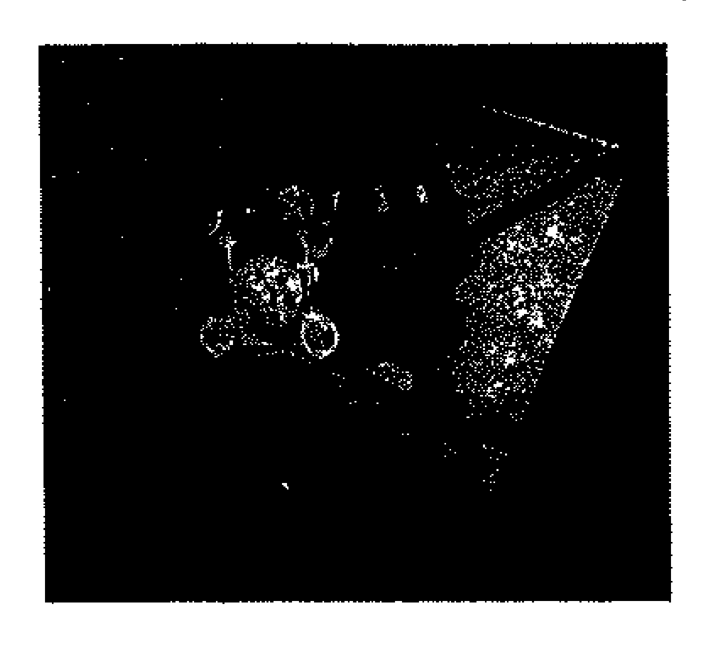
(١) اختبار الحجرة المائلة والكرسى المائل

:Tilting - Room - Tilting - Chair Test (TRTC)

ويُطلب من الفرد فى هذا الاختبار أن يُعدِّل من وضع الكرسى الذى يجلس عليه إلى الوضع الرأسى عندما يلمح أن الحجرة المائلة لا تحافظ على جسمه فى وضع متوازن.



ويوضح شكل (١٣-١٠) التالي هذا الاختبار :



شكل (١٣-١٠) يوضح اختبار الحجرة المائلة والكرسي المائل

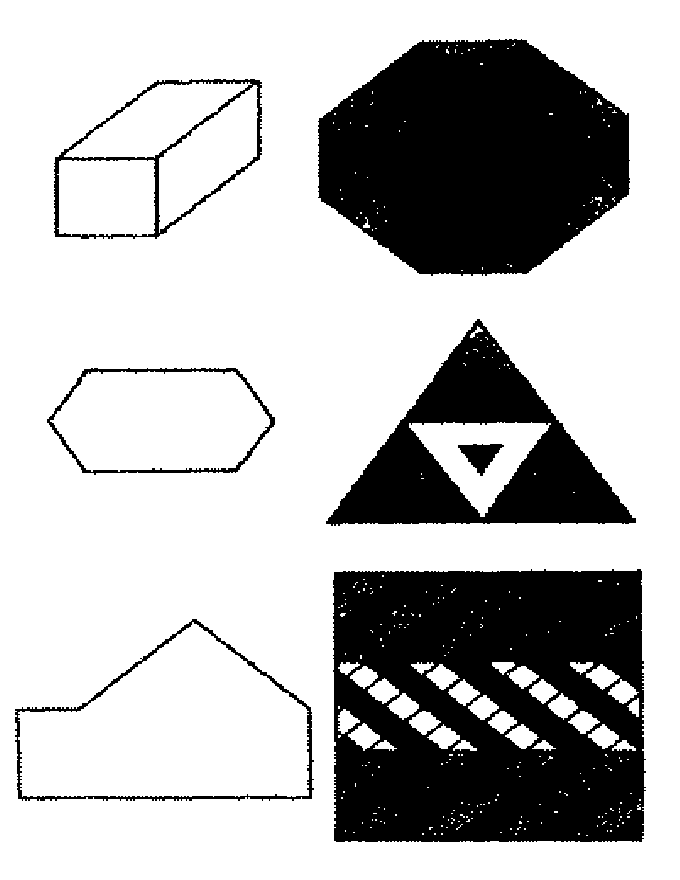
(٢) اختبار المؤشر والإطار Rod and Frame Test (RFT) :

ويُطلب من الفرد تعديل مؤشر مضيئ في وضع رأسي عندما يكون المرجع البصري الوحيد هو الإطار الذي يمكن أن يميل بزوايا معينة. وترجع صعوبة هذا الاختبار إلى أن الحجرة تكون مظلمة إظلامًا تامًا، والمربع المضيئ المحيط بالمؤشر يتم تدويره بزوايا مختلفة في الوقت الذي يعدِّل فيه الفرد أو يعيد تعديل المؤشر . ويُفترض أن الفرد لا يجب أن يُشتت انتباهه للأوضاع المختلفة للمؤشر، وإنما يُحدد الوضع الرأسي للمؤشر بإحساسه الجسماني الداخلي فقط، فالأفراد المعتمدون أو الأقل تمايزًا يستجيبون للتغيرات التي تحدث في الوسط التجريبي بدرجة أكبر من الأفراد المستقلين.

(٣) اختبار الأشكال المتضمنة Embedd Figures Test (EFT) :

ويُطلب من الفرد تحديد شكل بسيط متضمَّن أو مطمور في شكل أكثر تعقيدًا معطى له، وتوجد صيغ مختلفة من هذا الاختبار تناسب الراشدين، والأطفال ، وأطفال ما قبل المدرسة.

ويوضح شكل (١٣ ـ١١) التالى هذا الاختبار:



شكل (١٣ـ١١) يوضح بعض مفردات اختبار الأشكال المتضمنة

وعلى الرغم من أن هذه الاختبارات إدراكية، إلا أنها تُستخدم فى تقييم خصائص الشخصية، فقد وُجد أن الفرد المستقل عن المجال فى الاختبارات الثلاثة يتسم بالاكتفاء الذاتى فى علاقاته الاجتماعية، ولديه وسائل دفاعية وضوابط قوية توجه أفعاله، كما يتسم بالفكر المرتفع، ويلجأ إلى نفسه إذا ما صادف اضطرابًا فى شخصيته بدرجة أكبر من الفرد المعتمد على المجال.

وقد أوضحت كثير من الدراسات أن هذا الأسلوب المعرفى يعد سمة مستقرة نسبيًا ومتسقة وتتميز بدرجة ما من العمومية، وترتبط كما ذكرنا بعدد من متغيرات الشخصية، مثل القيادة، والمسايرة الاجتماعية (Witkin et al., 1975).

ومن الجدير بالذكر أن هناك أنواعًا أخرى من الأساليب المعرفية، مثل أسلوب الجمود ـ المرونة، والأسلوب التحليلى ـ التركيبى، وأسلوب التأمل ـ الاندفاع،

وأسلوب التمايز التصوري وغيرها؛ لذلك تتعدد أدوات قياس هذه الأساليب. فاختبار الأسلوب التصوري The Conceptual Style Test هو الذى أعده كاجسن وزملاؤه (Kagen et al. , 1963) يطلب فيه من الطفل انتقاء صورتين متفقتين من بين مجموعة معطاة من الصور وتوضيح سبب اتفاقهما، ويميز هذا الاختبار بين الأطفال التحليليين والأطفال التركيبيين.

كما أن اختبار سجل مفهوم الدور Role Construct Repretory Test (RCRT) يعد من أساليب التقييم التى نالت اهتمامًا ملحوظًا ، ويرتبط بالأساليب المعرفية، وقد استند بناؤه إلى نظرية كيلى Kelly فى الشخصية المتعلقة بالمفهوم الشخصى Personal Construct Theory التى تهتم بوصف الحالات الفردية ، حيث إن الفرد هو الذى ينظم ويؤسس عالمه بطريقته الخاصة . ويتكون الاختبار من عدة صيغ تشترك جميعها فى أنها تتطلب من الفرد انتقاء أسلوبه الخاص فى المقارنة بين الناس بدلا من استخدام النظام التصورى لمن قام ببناء الاختبار. فالاختبار عبارة عن قائمة تشتمل على ٢٤ عنوانًا لأدوار متعددة مثل:

«المعلم الذى تُفضله» ، «أكثر شخص ناجح تعرفه»، «أقرب صديق إليك» ، وهكذا. ويكتب الفرد اسم الشخص المناظر لكل من هذه الأدوار على بطاقة منفصلة، ثم يقوم الفاحص الكلينيكى باختيار ثلاث بطاقات، ويطلب من الفرد أن يوضح كيفية اتفاق اثنين منها واختلافهما عن الثالثة، إذ ربما يقول أن معلمه مخلص، وصديقه عطوف ، وهكذا. ويستمر الفاحص إلى أن تنتهى جميع البطاقات، ثم يقوم بدراسة النتائج لتحديد النظام المفهومى Construct System للفرد، والأبعاد التى تحدد تفكيره فيما يتعلق بالآخرين ، وبذلك يمكن توجيهه فى جلسات العلاج النفسى إلى أبعاد أخرى يحتاج إلى تعلمها لكى يغير من نظرته إلى الآخرين. وقد طوّر كيلى هذا الأسلوب بعد ذلك ليكون أكثر عمومية،حيث أوضح مفهوم التعقد المعرفى Cognitive Complexity الذى يشير إلى العدد الكلى للمفاهيم أو الأبعاد المختلفة التى يستخدمها الفرد فى تفكيره أو نظرته نحو الآخرين، ويوجد اتفاق عام على عدد المفاهيم الشخصية Personal Constructs التى يمكن أن تُحدد موقع الآخرين بالنسبة لها، وتتراوح بين البساطة المعرفية والتعقد المعرفى (Kelly , 1955 , 1958).

وكذلك توجد مقاييس تتعلق بالضبط المعرفى Cognitive Control، أو ما أطلق عليه روتر (Rotter ,1972) محل الضبط (الداخلى) مقابل محل الضبط (الخارجى) للتعزيز Locus of Control . ويشتمل مقياس روتر على أزواج من الفقرات ، تتعلق

إحدى فقرات كل زوج بتوقعات الفرد من حيث إنه يستطيع التحكم فى أمور حياته أو ضبطها، والأخرى تتعلق بتوقعاته أن القوى الخارجية هى التى تتحكم فى أموره، ومثال ذلك:

« يعتمد نجاح الفرد على ما يبذله من جهد» ، «إن النجاح يعتمد على الحظ».

وقد نال هذا المتغير المعرفى اهتمامًا ملحوظًا ، حيث أجريت حوله كثير من الدراسات، وأوضح بعضها أنه يرتبط ارتباطًا دالاً بمتغيرات أخرى للشخصية (Getter & Weiss , 1968; Stephens & Delys , 1973).

تقييم الأساليب الموضوعية :

تهتم الأساليب الموضوعية فى تقييم الشخصية بجمع بيانات عن الفروق الفردية بطريقة منظمة، وتصمم المقاييس المتعلقة بها بحيث تجعل الفرد يُركز على مهمة معينة ويكملها بسرعة أو فاعلية، وبملاحظة أو تسجيل بعض مظاهر أدائه الذى يرتبط ارتباطًا غير مباشر بالمهمة المعينة، يحاول الفاحص الحصول على معلومات عن نمط الوظائف المميزة له، وبذلك يمكن التغلب على الموقف الدفاعى للفرد عند تقييم شخصيته بطريقة مباشرة.

غير أن هذه المقاييس تتعلق بنمط معين من الأداء يتحدد بالإجراءات التى اتُّبعت فى قياسه. فإذا استُخدم المقياس أو الاختبار كمتغير تابع فى تجربة سيكلوجية، فإن النتائج يكون لها أهمية نظرية على الرغم من عدم إمكانية التوصل إلى تفسيرات تتسم بالعمومية، وذلك لأن درجات الاختبار تفتقر إلى أساس منطقى ، كما هو الحال عندما يُستخدم مقياس لنشاط فسيولوجى معين كمؤشر للعُصابية، مما يتطلب التحقق من صدق هذا المؤشر بدرجة عالية من الثقة. غير أن كثيرًا من الأساليب الموضوعية غير المباشرة تفتقر إلى دلالات الصدق المرجوة فى تقييم الشخصية، وتتطلب عملية جمع البيانات إجراءات تفصيلية مكلفة.

ومع هذا فإن هذه الأساليب لا تتأثر استجابات الأفراد لها بالاتجاهات العقلية Response sets التى تلعب دوراً واضحًا فى استبيانات الشخصية، ويمكن استخدامها كجزء من عملية التقييم الشامل لشخصية الفرد، على الرغم من أنها تتطلب مزيدًا من البحوث لتعرُّف مدى فائدتها فى التقييم الموضوعى للشخصية.

تعقيب عام على طرق وأساليب قياس الشخصية :

يتضح مما سبق تعدد أساليب قياس الشخصية، وتباين المداخل التى تستند إليها سواء فى بناء الاختبارات والمقاييس، أو فى تقدير الدرجات وتفسيرها، ومن بين هذه المداخل: المدخل المنطقى ، والمدخل السيكومترى، والمدخل الإمبيريقى، والمدخل التجريبى، والمدخل الكلينيكى التشخيصى، والمدخل الكلينيكى الفينومونولوجى. وأحد الفروق المهمة بين هذه المداخل الرئيسة فى تقييم الشخصية يتعلق بطريقة الحصول على البيانات وكيفية معالجتها.

فالأساليب الموضوعية السلوكية تعالج البيانات كعينة من سلوك الفرد، أما الأساليب الإسقاطية فتعتبر استجابات الفرد بمثابة مؤشر لدوافعه الداخلية وصراعاته ومشاعره، ومعظم الأخصائيين الذين يستخدمون هذه الأساليب تأثروا بدرجة كبيرة بنظرية فرويد والفكر المتعلق بالتحليل النفسى، بينما نجد أن الأساليب السيكومترية تعتمد فى تفسير درجات استبيانات الشخصية على الدرجات المعيارية بأنواعها المختلفة والصفحات النفسية (البروفيل) للاختبار.

- غير أنه لا يوجد اتفاق بين علماء النفس على كيفية ضم نتائج الملاحظة أو المقابلة مع درجات هذه الاستبيانات للتوصل إلى قرارات تتعلق بالأفراد أو المرضى أو التنبؤ بسلوكهم المستقبلى، وقد أدى ذلك إلى التساؤل حول ما إذا كان من الأفضل ضم البيانات بطريقة ذاتية أو كلينيكية أم بطريقة آلية أو إحصائية. فمزايا وعيوب الأساليب والبيانات الكلينيكية Clinical Methods فى مقابل الأساليب والبيانات الإحصائية Actuarial Methods لا تزال مثار جدل حاد ونقاش مستمر. وقد برزت هذه المشكلة بصورة أوضح فى كتاب ميهل (Meehl, 1954) «التنبؤ الكلينيكى فى مقابل التنبؤ الإحصائى» ، حيث ناقش مشكلة استخدام بيانات اختبارات ومقاييس الشخصية فى التنبؤ وأوضح أن التنبؤ الكلينيكى يتضمن صياغة فروض خاصة بتكوين شخصية الفرد أو ديناميتها، واستخدام هذه الفروض فى التنبؤ. ويمكن أن يستند التنبؤ إلى الحدس أو الحكم الشخصى، ولا يعتمد على العلاقات الإحصائية. فقد يعتمد الكلينيكى اعتمادا تامًا على الحدس Intuition، وهنا ربما لا يستطيع تحديد أساس تنبؤه، أو قد يستخدم المبادئ العامة المُستمدة من إحدى نظريات الشخصية ويُطبقها على الحالات الفردية.

أما فى التنبؤ الإحصائى فإن الباحث يُحدد العلاقة بين الدرجات المختلفة فى الاستبيانات أو الاختبارات والسلوك المحك المتنبأ به، ويُفسِّر الدرجة التى يحصل عليها الفرد بالرجوع إلى جداول المعايير، ولا يعتمد هذا التنبؤ على نظرية معينة، وإنما يقوم على أساس اعتبارات إحصائية فقط.



ومنذ نشر كتاب ميهل انقسم الأخصائيون الكلينيكيون إلى فريقين يتبنّى كل منهما إحدى وجهتي النظر، وحاول كل منهما التوصل إلى أدلة تؤيد وجهة النظر التي يتبناها وقد قام ميهل بمراجعة عدد كبير من الدراسات التي استخدم بعضها المدخل الاحصائي، واستخدم البعض الآخر المدخل الكلينيكي، وتوصل إلى أن أكثر من ثلثي الدراسات أكدت تفوق المدخل الإحصائي على المدخل الكلينيكي في التنبؤ الكلينيكي. أما بقية الدراسات فقد أوضحت أن المدخل الإحصائي يتساوى معه في التنبؤ. كما أكد ميهل أن المدخل الإحصائي يستغرق استخدامه وقتا وجهدا وكلفة أقل من المدخل الكلينيكي، وقصد بذلك أن الأخصائي الكلينيكي يقضي وقتا طويلاً في عمل يمكن أداؤه بدرجة أكثر فاعلية عن طريق شخص أقل كفاءة، أو باستخدام الآلة استناداً إلى الأساليب الإحصائية. ولكن ينبغي ملاحظة أن مناقشة ميهل اقتصرت على التنبؤ ولم تتعرض لنوع البيانات المستخدمة في التنبؤ. فالتمييز بين نوعي التنبؤ اقتصر على كيفية ضم البيانات التي يتم جمعها لأغراض التنبؤ. وبالطبع لا يعكس الانقسام بين الفريقين الأساليب المستخدمة في التنبؤ فقط، وإنما يعكس أيضا اختلافا واضحا في نظريات الشخصية وطرق التقدير التي يتبناها كل فريق (Meehl, 1959).

وعلى الرغم من ذلك فإنه يمكن استخدام كل من المدخلين في التفسير إذا علمنا جوانب القوة والضعف في كل من أسلوبي القياس أو التقدير، وفي كل من نوعي التنبؤ. فمعظم الكلينيكيين يستخدمون عدداً من الأساليب الإحصائية التي تساعدهم في اتخاذ القرارات. إذ ربما يلجأ أحدهم إلى تحويل الدرجات الخام إلى درجات معيارية، أو الرجوع إلى جداول معايير، أو رسم صفحة نفسية (بروفيل)، أو استخدام معادلات الانحدار لاتخاذ قرار معين، هذا إلى جانب التأمل فيما وراء البيانات المتعلقة بالسلوك وجوانب الاختبار.

لذلك فقد أوضح كرونباك (Cronbach, 1960) أن لكل مدخل في الحصول على البيانات المتعلقة بالشخصية وتفسيرها أغراض مناسبة. فإذا واجه أخصائي القياس النفسي مشكلة كلينيكية تتطلب الفهم وليس التقويم البسيط فقط، مثل تحديد أسباب السلوك العدواني لدى طفل، فإنه لا يستطيع الإجابة عن ذلك، وإذا حاول الأخصائي الكلينيكي تعرف أسباب ذلك دون الاستناد إلى نظرية معينة مناسبة، فإنه ربما تكون تفسيراته غير صحيحة، فأيٌّ منهما يمكن أن يتفوق على الآخر إذا عمل في مجاله الصحيح.

استخدامات مقاييس الشخصية :

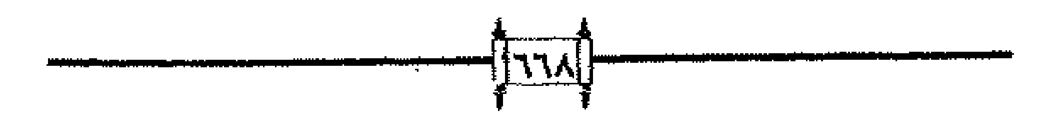
تُستخدم مقاييس الشخصية في غرضين رئيسيين، أحدهما نظرى بحثى، والآخر عملى تطبيقى، فالغرض النظرى البحثى يُمكّننا من تحديد متغيرات الشخصية وتكويناتها المرضية عن طريق قياس هذه المتغيرات ومحاولة الربط بينها وبين أنماط السلوك الملاحظ، أما الغرض العملى التطبيقى فيُمكّننا من اتخاذ قرارات والتنبؤ بسلوك الأفراد فى مواقف عملية متنوعة. فكثير من المشتغلين بعلم النفس فى مختلف المجالات التطبيقية، أو مجالات البحث يهتمون اهتمامًا خاصًا بقياس الشخصية، ومن بين هذه المجالات: التشخيص، والإرشاد والعلاج النفسى، وانتقاء الأفراد، وبحوث الشخصية وبناء النظريات، وكذلك فى تقويم البرامج . وسوف نوضح فيما يلى بإيجاز كلا من هذه الاستخدامات:

التشخيص: تُستخدم مقاييس الشخصية فى أغراض التشخيص الكلينيكى فى المستشفيات والعيادات النفسية، وذلك لتحديد مدى وطبيعة الاضطراب فى الشخصية الذى يُعانى منه المريض، واقتراح العلاج المناسب. وهذا يتطلب معرفة بنية وتنظيم الشخصية الذى يُؤدى إلى ما يعانيه المريض من اضطراب.

وقد عمل علماء النفس على مراجعة مفهوم التشخيص فى مجال الشخصية فى علم النفس والتربية، حيث إن النموذج الطبى Medical Model الذى يؤكد تحديد طبيعة المرض من تعرُّف أعراضه أصبح غير مقبول من جانب كثير من هؤلاء العلماء. فهذا النموذج لا يُمكّن الأخصائى من التوصل إلى تشخيص دون معرفة ظروف الفرد، ولا يكون هناك اتفاق بين مجموعة من الأخصائيين حول مُسببات العرض المرضى، وعدم اتساق تشخيصاتهم، ومعارضة بعض علماء النفس فكرة تصنيف الأنماط المرضية، مثل النمط العُصابى، والنمط الذُهانى، واقتراحهم محاولة فهم القوى والصراعات التى تدفع هؤلاء الأفراد إلى أن يسلكوا بطريقة معينة مميزة.

وقد اقتُرح النموذج السلوكى بديلاً عن النموذج الطبى، حيث يهتم بتعديل السلوك أكثر من اهتمامه بتشخيصه، أى محاولة التخلص من السلوك غير المرغوب فيه وتعزيز السلوك المرغوب.

الإرشاد والعلاج النفسى: يهتم الإرشاد والعلاج النفسى بعلاج اضطرابات الشخصية بأساليب سيكلوجية، فكلاهما يعتمد على التفاعل اللفظى بين المريض



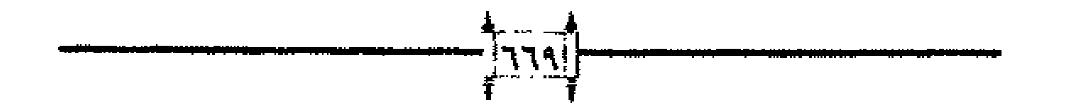
والأخصائى النفسى بغرض مساعدة المريض على تغيير سلوكه لكى يتيسَّر له التكيف، لذلك تُستخدم مقاييس الشخصية قبل وأثناء وبعد العلاج، بحيث تُصاحب التشخيص الكلينيكى، وذلك بهدف تقدير التغيُّر الذى ينتج عن أسلوب العلاج.

وقد نشأ علم الإرشاد النفسى كفرع منفصل يهتم بدرجة أساسية بالإرشاد التربوى والمهنى. فأخصائى الإرشاد النفسى يقوم بتوجيه الفرد إلى مجالات الدراسة أو المهن أو الوظائف التى تتناسب مع استعداداته وقدراته وميوله. غير أنه سرعان ما أدرك هؤلاء الأخصائيون أنه لا يمكن تجزئة الفرد إلى مجموعة من الاستعدادات والميول المنفصلة دون النظر إلى شخصيته كنظام متكامل؛ لذلك اتجهوا إلى تطبيق كثير من اختبارات ومقاييس الشخصية للتوصل إلى هذه الصورة المتكاملة.

أما علم النفس الكلينيكى فقد نشأ مرتبطًا ارتباطًا مباشرًا بمواقف الصحة النفسية، وتُصمم الأساليب التى يستخدمها الأخصائى الكلينيكى بحيث يمكن الإفادة منها فى تشخيص الاضطرابات النفسية. ولكن نظرًا لأنه لا يمكن فهم شخصية الأفراد فى ضوء أعراض الأمراض النفسية ، بل يجب التعرف على الجوانب الإيجابية فى شخصية هؤلاء الأفراد، فقد استُخدم مفهومى الإرشاد النفسى والعلاج النفسى ليحل كل منهما محل الآخر (Missick, 1961).

انتقاء الأفراد: تُستخدم مقاييس الشخصية وبخاصة أساليب الملاحظة والمقابلة واختبارات المواقف والاستبيانات، فى انتقاء الأفراد وبخاصة فى المجالات الصناعية والعسكرية . كما تُستخدم فى ترقية الأفراد، أو قبولهم فى برامج تدريبية معينة لزيادة فاعلية أدائهم. وتستخدم محكات متعددة، مثل تقديرات المشرفين على العمل، والإنتاجية، والنجاح فى مؤسسات التدريب، والمُعدَّل العام للطلاب. والمشكلة الرئيسة هنا هى ما إذا كان المتقدمون يتميزون بسمات شخصية معينة يتطلبها العمل المرشحون له، وهذه السمات يمكن تحديدها باستخدام مقاييس الشخصية. كما أن هناك صعوبات تتعلق بعدم ثبات درجات أو تقديرات هذه المحكات، وكذلك صعوبات تتعلق بطبيعة سمات الشخصية المسهمة فى النجاح فى المهن المختلفة.

وتُستخدم أحيانًا الاختبارات الإسقاطية فى أغراض الانتقاء، ولكنها تُؤدى إلى نتائج غير مرضية.



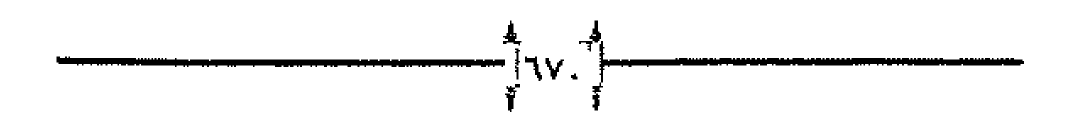
بحوث الشخصية وبناء النظريات :

تُستخدم مقاييس الشخصية أيضًا في التحقق من فرضيات تتعلق بطبيعة سمات الشخصية، وكيفية تنظيم هذه السمات ، وكيف تُستخدم في أغراض متعددة، وتلافي إساءة استخدامها. كما تستخدم في التحقق من كفاية طرق القياس ذاتها حيث تُصمم البحوث للكشف عما إذا كانت أداة قياس معينة تقيس سمات شخصية بطريقة متسقة، وما إذا كانت السمات التي تقيسها هي ما يود الباحث قياسها. فربما تدعو حاجة مؤسسة صناعية معينة أو مستشفى للأمراض النفسية أو العقلية إلى بناء مقياس جديد للشخصية لأغراض معينة، فهنا يقوم فريق البحث بدراسة مدى تحقيق المقياس للهدف الذي أعد من أجله.

ويمكن أيضًا تحديد العلاقات القائمة بين تكوينات فرضية، مثل الشخصية ، والذكاء، والميول، والاتجاهات، والتحصيل باستخدام أساليب التحليل العاملي لتوضيح معاني هذه التكوينات الفرضية واستخدام مقاييسها استخدامًا فاعلا . وكذلك يمكن بناء نظام منطقي يمكن فيه قياس هذه التكوينات الفرضية التي يمكن أن نستدل منها على العلاقات القائمة بين الأنماط السلوكية.

ونظراً لأن النظريات المستحدثة في مجال الشخصية ربما تتطلب أساليب أو مداخل جديدة في القياس تعتمد على أساليب النمذجة Computer Models of Personality، فإن استخدام الحاسوب في محاكاة الشخصية الإنسانية Computer Simulation يعد من الطرق المعاصرة في دراسة الشخصية، والتوصل إلى فروض مهمة عن هذه الظاهرة. فبدلا من تطبيق اختبارات ومقاييس الشخصية على الإنسان ربما يمكن اختبار الحاسوب بدلاً من الإنسان وفق برنامج معين، لتحديد مجالات التشابه والاختلاف بينهما مما يسهم في فهم طبيعة الشخصية الإنسانية فهمًا أفضل. ويمكن استخدام الحاسوب في إجراء مقابلات شخصية جماعية **Mass Interview** بغرض التعرف على الأفراد الذين ربما يحتاجون إلى علاج نفسي، كما يمكن استخدامه ـ كما سبق أن أوضحنا ـ في تفسير بيانات الصفحات النفسية (البروفيل) المستمدة من استبيانات الشخصية كما في استبيان مينيسوتا متعدد الأوجه **(MMPD)** وغيره من الاستبيانات.

وتُستخدم مقاييس الشخصية أيضًا في المواقف العملية التي تتضمن قياس أثر برنامج تدريبي معين مثلاً على الشخصية. وهنا تكون عينة البحث مجموعة من الأفراد المتطوعين للاشتراك في التجربة. فمثلا يمكن إجراء تجربة عن أثر الحرمان من الطعام لمدة طويلة على شخصية مجموعة من الأفراد.



تقويم البرامج:

على الرغم من أن مقاييس الشخصية ربما تكون محدودة الفائدة في تقويم البرامج التربوية والتدريبية، إلا أنها أحيانا تفيد في تعرف مدى فاعلية هذه البرامج في تعديل شخصية الأفراد. إذ يمكن أن تُطبق هذه المقاييس قبل بدء برنامج معين، وعقب الانتهاء منه لتعرف التغيرات التي حدثت نتيجة للبرنامج، وربما ترجع قلة استخدام مقاييس الشخصية في تقويم البرامج إلى ما سبق أن ذكرناه في هذا الفصل من تحيز الاستجابات، وصعوبة الحصول على محكات مناسبة ومتسقة، ونوعية الفقرات أو المهام المتعلقة بموقف معين دون سواه، وغير ذلك من العوامل التي تعوق الاستخدام الفاعل لهذه المقاييس في هذا الغرض.

كما أن تطبيق مقاييس الشخصية في مواقف تجريبية ربما يتنافى وحقوق الإنسان وخصوصياته إذا ما طُبقت عليه دون موافقته. وبعض هذه المقاييس يكون غامض الهدف، أو ربما يتعمَّد المُجرِّب إخفاء الهدف عن المشاركين في التجربة مما يتنافى والشروط الأخلاقية لاستخدام الاختبارات.

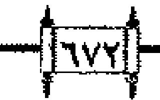
لذلك ينبغي مراعاة شروط وضوابط تطبيق الاختبارات والمقاييس التي أوضحناها بمزيد من التفصيل في الفصل الثاني.

خلاصة

عرضنا في هذا الفصل الأساليب والمقاييس المختلفة الشائعة الاستخدام في قياس الشخصية، وهي: الاستبيانات، والاختبارات الإسقاطية، والملاحظات وموازين التقدير، والمقابلة، واختبارات الأداء، والأنشطة الفسيولوجية، واختبارات الإدراك. وقد اتضح لنا مدى تعقد مفهوم الشخصية واختلاف تعريفاتها، حيث إن هناك عدم اتفاق بين علماء النفس في تصورهم لبنية أو مكونات الشخصية، وذلك لاختلاف الأطر النظرية التي يتبناها كل فريق منهم. وقد أدى هذا إلى عديد من المشكلات في قياس الشخصية، وذلك لاختلاف المبادئ المنظِّمة المستخدمة في التوصل إلى معلومات عن شخصية الأفراد، فبعضهم يستند إلى نظريات التحليل النفسي أو إلى المدخل الفينومولوجي، وبعضهم يستند إلى مدخل السمات التي يتم التوصل إليها باستخدام أساليب التحليل العاملي، وبعضهم يستند إلى الأساليب الإمبيريقية، في حين يستند فريق آخر إلى الأساليب التجريبية أو الموضوعية ـ السلوكية.

ونظرًا لتعدد هذه التوجهات والأساليب فقد اقتُرح تصنيف اختبارات ومقاييس الشخصية من حيث مدى تحديد بنيتها، ووضوح هدفها، وطريقة إجابتها. وحاولنا الربط بين نظريات الشخصية وطرق القياس والتقييم المختلفة. فنظريات الشخصية تؤدي إلى أنواع مختلفة من البيانات وتتأثر بها، كما أن طرق وأساليب قياس أو تقييم الشخصية تعتمد على فروض معينة تتعلق بطبيعة الشخصية وخطوات وإجراءات عملية القياس. فاستبيانات الشخصية تكون محددة البنية وواضحة الهدف، وتستند في بنائها إلى أساليب التحليل العاملي، مثل استبيان العوامل الستة عشر لكاتل (16 PF) أو استبيان مينيسوتا متعدد الأوجه (MMPI)، وغيرها من الاستبيانات التي تعتمد على التقرير الذاتي للفرد. أما الاختبارات الإسقاطية فتكون غير محددة البنية، وغامضة الهدف، حيث تشتمل على مثيرات غامضة مثل بقع الحبر، أو الصور التي تدور حول موضوع، أو الجمل الناقصة، أو الرسوم، أو اللوحات، أو اللعب وغير ذلك، ومن أهم هذه الاختبارات اختبار رورشاخ، واختبار تفهم الموضوع حيث يُعبِّر الفرد عما يراه ويُسقط مشاعره ودوافعه وعواطفه على هذه المثيرات الغامضة.

ويمكن تصنيفها في أساليب التداعي، والأساليب البنائية، وأساليب التكملة، وأساليب الاختيار أو الترتيب، والأساليب التعبيرية. وتُستخدم أساليب الملاحظة

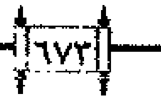


والتقدير، والمقابلة بكثرة فى تقييم جوانـب معينة من الشخصية، وبخاصة فى المواقف الكلينيكية، وفى أغراض انتقاء الأفراد، وتتفـاوت فى مدى تحديد بنيتها ووضوح هدفها وأساليب بنائها. وتتعدد أنواع مـوازين التقـدير التى يمكـن استخـدامها فى أساليب الملاحظة المقننة، كما تُستخدم أحيانًا فى المقابلة.

وتُوجد مجمـوعة أخرى من الأساليب التى تنـاسب المجالات البحثيـة، وتتعلق بالأنشطة الفسـيولوجية واختبـارات الإدراك التى تُعد حلقة وصل بين مـقاييس القدرات ومقاييس الشخصيـة. وتتعلق اختبارات الإدراك بالأساليب المعرفيـة، ومن بينها اختبار الحجرة المائلة والكرسى المائل، واختبار المؤشر والإطار، واختبار الأشكال المتضمنة التى أعدَّها ويتكن وزملاؤه لقياس الاستقلال أو الاعتماد الإدراكى على المجال.

وتختلف مقاييس الشخصية اختلافًـا ملحوظًا فى ثبات درجاتها وصدقها، ويرجع ذلك إلى اخـتلاف مصـادر الأخطاء المنتظـمة وغيـر المنتظمة التى تؤثر فى الـموقف الاختبارى وفى درجات المـقياس. فاستبيانات الـشخصيـة تتأثر بتـحيز الاستـجابات والاتجاهات العقلية للفرد وغير ذلك من الأخطـاء. والاختبارات الإسقاطية تتأثر بذاتية القائم بتـطبيق الاختبار وتـصحيحه وتفسير نتائجه، كـما أنه لا توجد أدلة كافـية تعزِّز استخـدام هذه الاختبـارات . وكذلك تفتقـر أساليب الملاحـظة والمقابلـة والأساليب الموضوعية ـ السلوكية إلى مؤشرات مرتـفعة للثبات والصدق. وعلى الرغم من محاولة القائـمين ببناء مـقاييس الشـخصية ربـط استجابات أو أداء الأفـراد فى هذه المقـاييس بسلوكهم أو أدائهم فى مواقف خارج نطاق المـوقف الاختبارى ، إلا أنهم يختلفون فى تحديد أهمية السلوك الـذى سترتبط به نتائج الاختبار، وفى مدى كفـاية الأدلة المتعلقة بهذه العلاقات.

ونتيجـة لهذه الاختلافات لاحظنا أن مـؤيدى كل طريقة يؤكدون مزايا طـريقتهم ويشيرون إلى قصور الطـرق الأخرى، وهذا ربما يبين لنا مدى تعقد قيـاس الشخصية؛ لذلك لا نستطيع القول بأن هناك أسلوبًا أمثل لقيـاس الشخصية، وإنما ينبغى النظر إلى كل من هذه الأساليب نظرة فاحصة ناقدة، وانتقاء ما يناسب الغرض من الاختبار، وفيم تُستخـدم نتائجه . فاختبـارات ومقاييس الشخصية تُستخدم فى العديـد من المجالات الكلينيكية، والإرشاد والـعلاج النفسى، وانتقاء الأفراد، وتقويم البرامـج، كما تُستخدم فى الأغراض البحثية والتحقق من صحة الفروض المتعلقة بشخصية الأفراد، والعلاقات بين متغـيرات الشخصيـة، والمتغيرات الأخرى الـمتعلقة بالذكـاء، والإدراك، والميول وغيرها.



وعلى الرغم من أن الأساليب المختلفة لقياس الشخصية تُقدّم معلومات ومؤشرات مهمة عن طبيعة وتنظيم الشخصية، إلا أن التفسيرات المستمدة من أى منها ينبغى أن تُضم إلى البيانات التى يتم جمعها باستخدام الأساليب الأخرى، وكذلك إلى البيانات المستمدة من اختبارات التحصيل، والذكاء، والاستعدادات الخاصة ، والميول والاتجاهات التى ناقشناها فى الفصول السابقة من أجل فهم أفضل لشخصية الفرد، والتنبؤ بسلوكه المستقبلى. فالشخصية الإنسانية تُعد نظامًا مفتوحًا Open System تتفاعل فيه المكونات المعرفية والوجدانية والمزاجية وغيرها، كما تتفاعل مع المتغيرات البيئية وتتأثر بها بدرجة كبيرة كما تؤثر فيها، مما يؤدى إلى تباين ملحوظ فى سلوك الأفراد ومدى تكيفهم مع أنفسهم ومع بيئتهم المحيطة بهم.



توجهات معاصرة فى القياس
والتقويم التربوى والنفسى



الفصل الرابع عشر

بعض التطورات النظرية المعاصرة

* مقدمة

أولا: نظرية الاستجابة للمفردة الاختبارية

* نظرية الاستجابة للمفردة الاختبارية من منظور تاريخى
* الأنواع الرئيسة لنماذج الاستجابة للمفردة الاختبارية
* الأساس المنطقى لنماذج الاستجابة للمفردة الاختبارية
* تصنيف نماذج الاستجابة للمفردة الاختبارية
* مفهوم المنحنى المميز للمفردة (ICC)
* النماذج الاستاتيكية اللوغاريتمية أحادية البُعد
* نموذج راش
* نموذج البارامترين
* نموذج البارامترات الثلاثة
* مزايا وعيوب نماذج الاستجابة للمفردة الاختبارية

ثانيًا: نظرية الثبات متعدد الأوجه

* بعض أوجه قصور المفهوم الكلاسيكى للثبات والصدق
* بعض المفاهيم الأساسية لنظرية إمكانية التعميم
* دراسات إمكانية التعميم
* دراسات القرارات
* تحديد النطاقات الشاملة فى دراسات إمكانية التعميم
* تحليل مكونات التباين فى دراسات إمكانية التعميم
* العلاقة بين نظرية إمكانية التعميم ونظرية الصدق
* صدق التكوين الفرضى من منظور نظرية إمكانية التعميم

ثالثًا: إسهام علم النفس المعرفى فى تطوير أساليب القياس والتقويم

* علاقة النظرية المعرفية بالعملية الاختبارية
* تصميم الاختبارات فى ضوء النظرية المعرفية
* خلاصة

مقدمة :

عرضنا في الأبواب الثلاثة السابقة الأسس النظرية وأهم الجوانب التطبيقية المعرفية والوجدانية والشخصية للقياس والتقويم التربوى والنفسى، ولاحظنا مدى تعدد وتنوع أدوات قياس هذه الجوانب، وثراء الأساليب التى يمكن استخدامها فى هذا المجال الحيوى الإنسانى، وتطبيقاتها فى مختلف مجالات تقويم الأفراد والبرامج التربوية والنفسية والتدريبية والمهنية. وسوف نتناول بإيجاز فى هذا الباب بعض أهم التطورات المعاصرة المتعلقة بالأسس النظرية، وأساليب بناء الاختبارات والمقاييس التربوية والنفسية وتحليل وتفسير البيانات المستمدة منها، وماترتب على ذلك من توجُّهات جديدة فى استخدامات هذه الاختبارات والمقاييس وتطبيقاتها المتنوعة.

فالذى يطلع على أدبيات القياس المعاصر يلاحظ بوضوح إسهام التطورات التكنولوجية إسهاما كبيرا فى تقدم علم القياس والتقويم التربوى والنفسى الذى أدى بدوره إلى إعادة النظر فى النظريات والنماذج السيكومترية المتعارف عليها، وتطوير نماذج ونظريات معاصرة للتغلب على كثير من المشكلات المتعلقة ببناء الاختبارات والمقاييس فى الجوانب المختلفة للسلوك الإنسانى، بما يجعلها أكثر مرونة وملاءمة لطبيعة الظواهر السلوكية.

وبالطبع يصعب تناول أو توضيح جميع جوانب القياس والتقويم التى شملها التطور فى هذا الباب، إذ إن كلا منها ينبغى أن يفرد له عدة مجلدات، ولكننا سوف نلقى الضوء على بعض أهم هذه التطورات التى تجعل الدارس والباحث والممارس أكثر استبصارا فى بنائه أو انتقائه أو استخدامه لأدوات القياس والتقويم التربوى والنفسى والاجتماعى، بحيث يمكنه الرجوع إلى المصادر المتخصصة التى تناولت هذه التطورات. وسوف نتناول فى هذا الفصل بعض أهم التطورات النظرية فى هذا المجال، ونوضح بعض القضايا والتوجهات المتعلقة بتطبيقاتها فى الفصل الخامس عشر.

ومن أهم التطورات المعاصرة التى سوف نلقى عليها الضوء فى هذا الفصل نظرية الاستجابة للمفردة الاختبارية Item Response Theory (IRT) ، ونظرية الثبات متعدد الأوجه Generalizability Theory، وبعض إسهامات علم النفس المعرفى فى تطوير أساليب القياس والتقويم. ويرجع اختيارنا لهذه الموضوعات إلى أهميتها واستخداماتها المتعددة فى مختلف المجالات التطبيقية فى وقتنا الحاضر.

أولا: نظرية الإستجابة للمفردة الإختبارية

Item Response Theory (IRT)

اعتمدت الاختبارات والمقاييس التى أوضحناها فى البابين الثانى والثالث فى أساليب بنائها وتحليل مفرداتها على أسس ومبادئ النظرية الكلاسيكية للاختبارات Classical Test Theory (CTT). وقد أوضحنا بعض هذه الأسس والمبادئ فى الباب الأول، وأوضح كل من جاليكسان (Gulliksen, 1950) فى كتابه نظرية الاختبارات العقلية Theory of Mental Measurement، وجيزلى (Ghiselli, 1964) فى كتابه نظرية القياس النفسى Theory of psychological Measurement ، وماجنسون (Magnusson, 1966) فى كتابه Test Theory، ونانلى (Nunnally, 1987) فى كتابه النظرية السيكومترية Psychometric Theory - أسس ومبادئ هذه النظرية بالتفصيل فى مجال الاختبارات مرجعية الجماعة أو المعيار (NRTs).

ويعتمد انتقاء مفردات هذه الاختبارات فى إطار هذه النظرية على نتائج بعض الإحصاءات المتعلقة بمايلى:

(١) تجانس مفردات الاختبار Item Homogeneity: ونقصد بذلك أن الأفراد الذين يجيبون إجابة صحيحة عن إحدى مفردات اختبار يميلون إلى أن يجيبوا إجابة صحيحة عن بقية مفرداته، وبذلك يحصل كل منهم على درجة مرتفعة فى الاختبار، ويحدث العكس بالنسبة للأفراد الذين يجيبون إجابة خطأ عن إحدى المفردات. ويُطلق على مقياس تجانس المفردات اسم «معامل التمييز Discrimination Index»، الذى يمكن الحصول على قيمته باستخدام معامل الارتباط الثنائى المتسلسل بين درجات الأفراد على المفردات ودرجاتهم الكلية فى الاختبار.

(٢) الاتساق الداخلى للاختبار Internal Consistency: ونقصد بذلك أن يكون احتمال إجابة فرد إجابة صحيحة عن إحدى مفردات اختبار دالة مطردة لموقع الفرد على متصل السمة أو القدرة التى يقيسها الاختبار. فإذا كان نمط إجابات الأفراد الذين يحتلون نفس الموقع على هذا المتصل عن جميع مفردات الاختبار متشابها، يمكن الحصول على معاملات الاتساق الداخلى باستخدام معامل كيودر - ريتشاردسون Kuder - Richardson الصورتين ٢٠، ٢١، أو معامل ألفا لكرونباك.

(٣) صعوبة مفردات الاختبار Item Difficulty: وتقدَّر بنسبة عدد الأفراد الذين يجيبون إجابة صحيحة عن مفردة معينة من مفردات الاختبار.



(٤) التشتت الأقصى لتوزيع درجات الاختبار Maximal Dispersion:

ونقصد بذلك أن توزيع درجات الأفراد على متصل السمة أو القدرة التى يقيسها الاختبار يتخذ شكل المنحنى الاعتدالى. لذلك يُعنى بانتقاء المفردات التى تتراوح قيم معامل صعوبتها بين ٠,٤٠ ، ٠,٧٠ لكى يكون انتشار الدرجات أكبر مايمكن. ويعد هذا الشرط قيدا تفرضه النظرية الكلاسيكية للاختبارات مرجعية الجماعة أو المعيار على توزيع الدرجات. وقد ناقشنا هذه الإحصاءات فى الفصول الأربعة الأخيرة من الباب الأول.

ولكن ما المشكلات المتعلقة بالاختبارات التى يتم بناؤها فى إطار هذه النظرية؟

إن المشكلة الرئيسة هى أن جميع الخصائص السيكومترية للاختبارات التى أشرنا إليها تعتمد على خصائص عينة الأفراد التى يُجرى عليها الاختبار، وعلى خصائص عينة المفردات التى يشتمل عليها الاختبار، مما يؤدى إلى نظام قياس نفسى وتربوى غير مستقر.

فمعاملات صعوبة المفردات تتذبذب وتتغير بتغير سمات أو قدرات أفراد العينة المراد اختبارها، وقياس سمات أو قدرات الأفراد تتذبذب وتتغير بتغير صعوبة مفردات الاختبار، وثبات الدرجات المستمدة من الاختبار تتذبذب وتتغير بتغير مستوى وانتشار سمات أو قدرات عينة الأفراد، والدرجات المستمدة من هذه الاختبارات لايكون لها معنى أو دلالة فى ذاتها، إذ يختلف معنى الدرجة باختلاف سهولة أو صعوبة مفردات الاختبار، وباختلاف ضيق أو اتساع مدى الاختبار. لذلك إذا أردنا المقارنة بين مستويى أداء فردين، أو مقارنة أداء الفرد نفسه فى موقفين مختلفين، فإننا يجب أن نستخدم الاختبار نفسه فى أى من الحالتين. ولو قمنا ببناء اختبارين يشتملان على نفس العدد من المفردات التى تقيس سمة أو قدرة معينة لدى مجموعة من الأفراد، فإن درجة كل فرد سوف تختلف باختلاف صعوبة مفردات كل من الاختبارين؛ لذلك يُقال أن الاختبارات التى تبنى بالطرق السيكومترية الكلاسيكية محدودة بعينة المفردات Item - bounded، وبأفراد العينة التى استخدمت فى تقنين الاختبار Sample - bounded. وهناك مجموعة أخرى من المشكلات فى هذا الشأن نوجزها فيما يلى:

(١) يُفترض أن الاختبار يقيس متغيرا أحادى البُعد، وربما لايكون ذلك صحيحا فى كثير من الأحيان.

(٢) يُفترض أن الدرجات التى يحصل عليها فرد فى مفردات الاختبار يمكن جمعها كما لو كانت تُمثل ميزانا خطيا Linear Scale، وأن المفردات المتعلقة بالمتغير



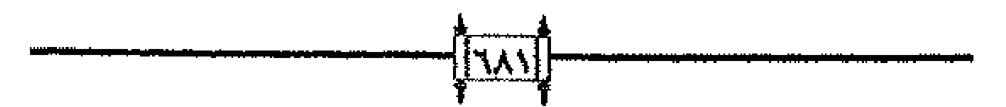
المراد قياسه تحمل المعنى نفسه لدى جميع المختبرين. والحقيقة أن هذا الميزان يكون منحنيا، فالفرق الثابت بين درجتين من درجات الاختبار يختلف معناه بالنسبة لدرجة السمة أو القدرة المقاسة بحسب موقع الدرجات على متصل السمة أو القدرة.

(٣) يُفترض أن درجات الاختبار التى تُمثل السمة أو القدرة المقاسة دالة خطية مُطردة Monotone Function، بمعنى أنه كلما زادت درجة الفرد فى الاختبار، دل ذلك على زيادة مقدار السمة أو القدرة لديه. غير أننا نلاحظ أن بعض الأفراد من ذوى القدرات المرتفعة يحصلون أحيانا على درجات منخفضة فى الاختبارات، وربما يحدث العكس بالنسبة لذوى القدرات المنخفضة.

(٤) تغيُّر تكوين الاختبار بمرور الزمن Test Monolithic Structure. ونعنى بذلك أنَّ تكوين ومعنى مفردات الاختبار يتغير من زمن إلى آخر بالنسبة لمجموعات الأفراد التى أُعِدّ لها الاختبار. فالظروف البيئية تتغير، والمواقف الاختبارية ليست دائما مقننة، وحذف أو تغيير أى مفردة من مفردات الاختبار يؤدى إلى تغير فى درجات الأفراد بطريقة يصعب التنبؤ بها، وبذلك يصبح الاختبار أداة قياس وحيدة. وعندما يتبين قصوره وعدم جدوى استخدامه بسبب الشك فى صدق البيانات المستمدة منه، وقدم معاييره؛ يُستبعد الاختبار ويحل محله اختبار مقنن آخر حتى تدور عليه الدورة وهكذا.

ولقد حاول علماء القياس التغلب على هذه المشكلات وغيرها، وتطوير نظم القياس النفسى والتربوى. فعملية القياس هى نتاج تفاعل بين الأشياء أو الخصائص المراد قياسها وبين أدوات القياس. ونتائج القياس يجب أن تكون مستقلة عن أداة القياس المستخدمة. ويرى علماء القياس المعاصرين (DeGruijter, 1984; Hambleton, 1978; Lord, 1980; Samejima, 1974; Wright, 1983) أن بناء نظام موضوعى لتحويل الملاحظات إلى قياسات يتطلب أن تتحرّر تداريج أدوات القياس من خصائص الأشياء التى تقيسها هذه الأدوات Object - Free Instrument Calibration وأن يتحرر قياس الخصائص أو الأشياء من خصائص أدوات القياس التى استخدمت فى هذا القياس Instrument - Free Object Measurement.

وهذا يتطلب أن يشترك كل من قياس الفرد Person Measurement وتعيير المفردات Item Calibration فى وحدة قياس ثابتة على مدى مُتصل المتغير. أى يكون تدريج هذا المتصل خطيا (مستقيما) ووحداته متساوية.



وبذلك يسمح هـذا النظام بتعميم الـقياسات خارج حدود الأداة التـى أُجرى بها القياس، كما يسمح بالموازنة بين خصائص الأشياء المراد قياسها بأدوات قياس متشابهة ولكن غير متطابـقة، ويمكننا من ضم أو تجزئة أدوات القياس بحيـث تناسب متطلبات قياس جديدة كما هو الحال فى القياس الفيـزيائى. فالنظـرية الكلاسيكـية للاختـبارات تؤدى إلى بناء اختبارات غير مرنة Traditional Fixed Tests ،بينما وجه علماء القياس جهودهم نحو بناء نظام قياس موضوعى يُركز على المفردة الاختبارية ذاتها Individual Item وليس على الاختبار ككل. وهذا النظام يسمح بإضافة مفردات أخرى إلى الاختبار أو حذف بعضها دون أن يتأثر الاختبـار ككل. وهذا يعنى بناء اختبارات مرنة Flexible Tests، أى يمكـن أن يكون الاختـبار قصـيرا أو مفـرداته كثـيرة، أو ضيق الـمدى أو متسعًا، أو يتميز بالصعوبة أو السهولة ومع هذا يظل نظام تعبيره أو تداريجه مستقرًا.

وقد أدت هذه الجـهود إلى تطوير نظـرية معاصرة فى الـقياس النفـسـى والتربوى أطلق عليها نظرية الاستجـابة للمفردة الاختبارية Item Response Theory (IRT) ، وكان يطلق علـيها من قبل نظرية السـمات الكامنة Latent Trait Theory (LTT) ، ويندرج تحتهـا مجموعة من النماذج تعرف بـاسم نماذج الاستجابة للمفـردة الاختبارية Item Response Modols.

فهذه النماذج تسمح بالقياس الموضوعى الذى أشرنا إليه، كما يمكن باستخدامها مواجهة كثير من مشكلات القياس النظرية والتطبـيقية التى عجزت عن مواجهتها النظرية الكلاسيكية.

نظرية الإستجابة للمفردة الإختبارية من منظور تاريخي:

يرجع الفضـل فى تقديم أسس نظرية الاستـجابة للمفردة الاخـتبارية للمهـتمين بالقياس النفـسى والتربوى إلى لورد (Lord, 1952, 1953) . وقد تطوَّرت هـذه النظرية ومايتعلق بها من نماذج سيكومتـرية منذ ذلك الوقت بهدوء وثبات، وظلت أعمال لورد Lord وغيره من علماء القياس فى هذا المجال حـتى السبعينيات من هذا القرن لايلتفت إليها إلا عدد محـدود من خبراء القياس فى الولايات المتحـدة الأمريكية وبعض الدول المتطورة. غير أنه حدث تقـدم سريع فى الحركة البحثية والتطبيقـية لهذه النظرية ابتداء من عـام ١٩٦٨، وهو العـام الذى نشـر فيـه لورد Lord كتابه «النظريـات الإحصائية لدرجـات الاختبـارات العقلية Statistical Theories of Mental Test Scores» (Lord, 1968)، الذى جـمع بين غـلافيـة لأول مرة أسـس النظريـة السيـكومتـرية الكلاسيكية، وأسس نظرية الاستجابة للمفردة الاختبارية.

وقد اهتم جورج راش G. Rasch وهو أحد علماء الرياضيات والإحصاء بجامعة كوبنهاجن بالدنمارك بالقياس النفسي منذ عام ١٩٤٥ ، وتوصل إلى نموذج سيكومترى يمكن أن يندرج تحت نظرية الإستجابة للمفردة الاختبارية، ويعرف بنموذج راش Rasch Model واشتغل بالإحصاء الرياضى والإحصاء الحيوى Biometrics منذ ذلك الوقت.

وقد قام راش بزيارة الولايات المتحدة الأمريكية عام ١٩٤٧ للمشاركة فى تأسيس مجتمع القياس الحيوى. وأثناء زيارته لجامعة شيكاغو قابل سافيج J. Savage وهو أحد علماء الإحصاء البارزين بهذه الجامعة حيث دعاه لتكرار زيارة الجامعة عام ١٩٦٠ لكى ينتهى من إعداد كتابه الرائد «النماذج الاحتمالية لبعض اختبارات الذكاء والتحصيل» ويلقى عدة محاضرات فى هذا المجال. وقدّم راش العديد من الأوراق منذ عام ١٩٦٠ حول القياس الموضوعى وبخاصة ورقته حول « القوانين العامة ومعنى القياس فى علم النفس» الذى قدمها فى مؤتمر الإحصاء الرياضى بجامعة كاليفورنيا عام ١٩٦١ ، وورقته «تحليل المفردات الذى يأخذ الفروق الفردية بعين الاعتبار» عام ١٩٦٦ . وقد التقى راش بعالم النفس بنجامين رايت B. Wright بجامعة شيكاغو عام ١٩٦٠ وقدّمه لهذا الفكر السيكومتري الجديد الذى يتبناه رايت wright بعد ذلك. وقدّم ورقة حوله عام ١٩٦٧ بعنوان «تغيير الاختبار وقياس الأفراد متحررا من العينة»، وتلى ذلك نشاط بحثى كبير قامت به جامعة شيكاغو بإشرافه استنادا إلى النموذج السيكومترى الذى توصّل إليه جورج راش مستقلا عن نماذج الاستجابة للمفردة الاختبارية الأخرى التى توصل إليها بيرنبوم Birnbaum، ولورد Lord فى الولايات المتحدة الأمريكية، والنماذج الأخرى التى توصل إليها بعض خبراء القياس فى استراليا ونيوزيلندة. غير أنه يرجع الفضل فى انتشار هذه النماذج، وبخاصة نموذج راش وتوسيع مجالات تطبيقاته واستخداماته إلى كل من رايت Wright، وهامبلتون Hambleton، وشوبين Choppin.

وتنال هذه النماذج فى وقتنا الحاضر اهتماما بحثيا كبيرا يمكن ملاحظته بوضوح فى أدبيات القياس النفسى والتربوى المعاصر. وتحاول هذه البحوث التغلب على العديد من المشكلات المتعلقة بتقدير بارامتراتها وتقييم حسن مطابقة البيانات لهذه النماذج، وإمكانية استخدامها عندما تكون المتغيرات التى تقيسها الاختبارات متعددة الأبعاد، وغير ذلك من القضايا السيكومترية والتطبيقية.

ويرجع سبب عدم انتشار مبادئ وأسس هذه النظرية المعاصرة انتشارا واسعا إلى عدة أسباب نذكر منها:

(١) تمثّل هذه النظرية فرعًا معقدًا من فروع نظرية الاختبارات، واستيعاب البحوث السيكومترية التي تُجرى في هذا المجال وتُنشَر في الدوريات العلمية المتخصصة يحتاج إلى قدر كبير من الرياضيات العالية، والإحصاء متعدد المتغيرات، وإحصاء بيز.

(٢) معظم خبراء القياس الذين تناولوا هذا الموضوع وجّهوا دراساتهم وبحوثهم لمنظّرى القياس، وليس للذين سيقومون بتطبيقه في بناء الاختبارات وتحليل وتفسير نتائجها، وفي غير ذلك من التطبيقات السيكلوجية والتربوية.

(٣) لم يكن هناك برامج حاسوب فعالة ومناسبة لتقدير بارامترات نماذج السمات الكامنة التي تتضمنها هذه النظرية.

(٤) صعوبة تحقّق بعض فروض النماذج المتعلقة بهذه النظرية في البيانات الفعلية المستمدة من الاختبارات العقلية والتحصيلية.

لذلك اتجه علماء القياس في الآونة الأخيرة إلى مواجهة هذه الصعوبات مواجهة إيجابية، حيث أجريت كثير من الدراسات لتحويل هذه النماذج من الجانب السيكومترى النظرى البحت إلى الجانب العملى التطبيقى وتبسيطها حتى يتمكن المهتمون بالتطبيق من استخدامها دون عناء. وأُعدّت كثير من برامج الحاسوب التى يمكن الاستعانة بها فى تقدير بارامترات هذه النماذج. (Mislevy & Bock, 1985; Wingersky, Barton & Lord, 1982; Wright & Mead, 1975)

ونتيجة لهذه التطورات اتجهت كثير من المؤسسات والهيئات الأمريكية المسئولة عن إعداد الاختبارات ونشرها إلى استخدام هذه النماذج فى تصميم وبناء مختلف أنواع الاختبارات والمقاييس العقلية والتحصيلية،وموازين التقدير،ومقاييس الاتجاهات.

كما استعانت بها إدارات التقويم التربوى بمختلف الولايات فى إعداد الاختبارات الصفية فى المواد الدراسية المختلفة. وحذت بعض دول العالم مثل كندا ، وانجلترا، واستراليا، ونيوزيلنده، حذو الولايات المتحدة الأمريكية فى هذا الشأن.

الأنواع الرئيسة لنماذج الاستجابة للمفردة الاختبارية:

تهدف نماذج الاستجابة للمفردة الاختبارية لتحديد علاقة بين أداء الفرد فى اختبار معين، وهو مايمكن ملاحظته ملاحظة مباشرة، وبين السمات أو القدرات التى تكمُن وراء هذا الأداء وتفسره. وبالطبع تختلف السمات الكامنة Latent Traits باختلاف مايقيسه الاختبار. فالسمة الكامنة التى تنطوى عليها استجابات الأفراد لاختبار



فهم معانى الكلمات مثلا تختلف عن تلك التى تنطوى عليها استجاباتهم لاختبار العمليات الحسابية، أو اختبار الإدراك المكانى. وتُعد هذه النماذج دوال رياضية احتمالية تختلف باختلاف عدد بارامتراتها أو مكوناتها.

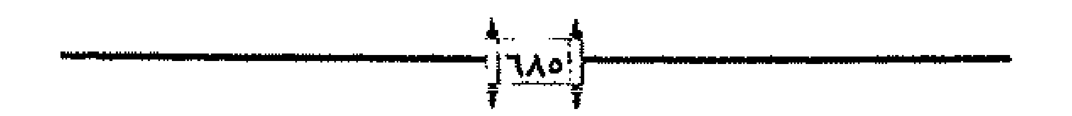
وتنقسم هذه النماذج إلى مجموعتين رئيستين كما يلى:

النماذج الاستاتيكية Static Models: وتهتم هذه النماذج بالقياس فى وقت معين، كما تهتم بتحديد العمليات التى ينطوى عليها الأداء فى الاختبارات التربوية والنفسية. فمثلا يمكن نمذجة الأداء فى اختبار للاستعداد الدراسى أو التحصيل بواسطة هذا النوع من النماذج. ومن أمثلة هذه النماذج نماذج راش، ونموذج لورد، ونموذج بيرنبوم وغيرها.

النماذج الديناميكية Dynamic Models: وتهتم هذه النماذج بمشكلة قياس التحسن أو التغير الذى يحدث فى السمات النامية المختلفة فى فترات زمنية متباعدة. وبعض هذه النماذج يعتبر التغير النمائى عملية تدريجية، أى أن التغير متصل، بينما يعتبر البعض الآخر أن هذا التغير انتقالى من حالة إلى أخرى، أى تغير منفصل، مثل الانتقال من حالة عدم التمكن إلى حالة التمكن فى الاختبارات مرجعية المحك. وتركزت غالبية هذه النماذج على الحالة الأولى، مثل نموذج بوك (Bock, 1972) الذى يتعلق بمنحنيات قياس التحصيل فى ضوء نموذج بيرنبوم الاستاتيكى، أما نموذج فيشر (Fischer, 1973) فقد اعتمد على نموذج راش.

ويرى أصحاب كلٍّ من نوعى النماذج أن النماذج الكلاسيكية فى بناء الاختبارات (CTM) تناسب قياس المتغيرات الناجمة عن ملاحظات أو أحكام منفردة Single Measurement، مثل تقدير المعلم لطالب فى سمة معينة أو قياس الطول أو الوزن، ولكن يصعب تطبيقها فى حالة المتغيرات الناجمة عن قياسات مركبة Composite Measurement من ملاحظات متعددة غير متكافئة، مثل المجموع الموزون للدرجات مجموعة من المفردات يشتمل عليها اختبار معين. فهذه المفردات تكون عادة متباينة فى صعوبتها وصدقها فى قياس السمة أو السمات التى يُفترض أنها تنطوى عليها. ونماذج الاستجابة للمفردة الاختبارية (IRM) تُطبق بدرجة أفضل فى هذه الحالة.

ولكن بعد تكوين «المتغير» فى الحالة الثانية بطريقة مناسبة يصبح متغيرا من الحالة الأولى، وهنا يمكن استخدام النماذج الكلاسيكية.



الأساس المنطقي لنماذج الاستجابة للمفردة الاختبارية:

يمكن إيجاز الفكرة الأساسية لنماذج الاستجابة للمفردة الاختبارية في أن هذه النماذج تحاول اشتقاق قياسات أو قيم تقديرية للسمة أو السمات التي تنطوي عليها مجموعة من الاستجابات لمجموعة من المفردات الاختبارية. وهذه القياسات تتميز بخصائص تفسيرية تتخطى حدود المجموع الموزون لدرجات الاختبار. أي أن هذه النماذج تفسر استجابة الفرد للمفردة الاختبارية التي تنطوي على السمة التي تقيسها هذه المفردة. وعادة يفترض أن السمة المقاسة هي قدرة معينة، أو خاصة من خصائص الفرد الذي يختبر بها. بحيث إنه توجد علاقة منتظمة بين مستويات السمة المقاسة لدى أفراد مختلفين وبين احتمالات الاستجابة الصحيحة لمفردات مختلفة.

لذلك فإن هذه النماذج تُعد نماذج إحتمالية Probabilistic Models، وتستند جميعها إلى دالة الترجيح اللوغاريتمي Logistic Function بدلاً من دالة الكثافة الاحتمالية الاعتدالية Normal Density Function التي تعتمد عليها النماذج الكلاسيكية التي استند إليها جميع الاختبارات والمقاييس التي أوضحناها في البابين الثاني والثالث. كما تستند إلى فروض تختلف عن فروض هذه النماذج الأخيرة سوف نوضحها بعد قليل.

تصنيف نماذج الاستجابة للمفردة الاختبارية:

تتعدد كل من النماذج الاستاتيكية والنماذج الديناميكية التي أشرنا إليها وفقاً للمتغيرات التي يتضمنها النموذج، والتي يرى أصحاب هذه النماذج أنها تلعب دوراً في الاستجابة للمفردة الاختبارية. غير أنه يمكن تصنيف هذه النماذج في ضوء محكات أربعة هي:

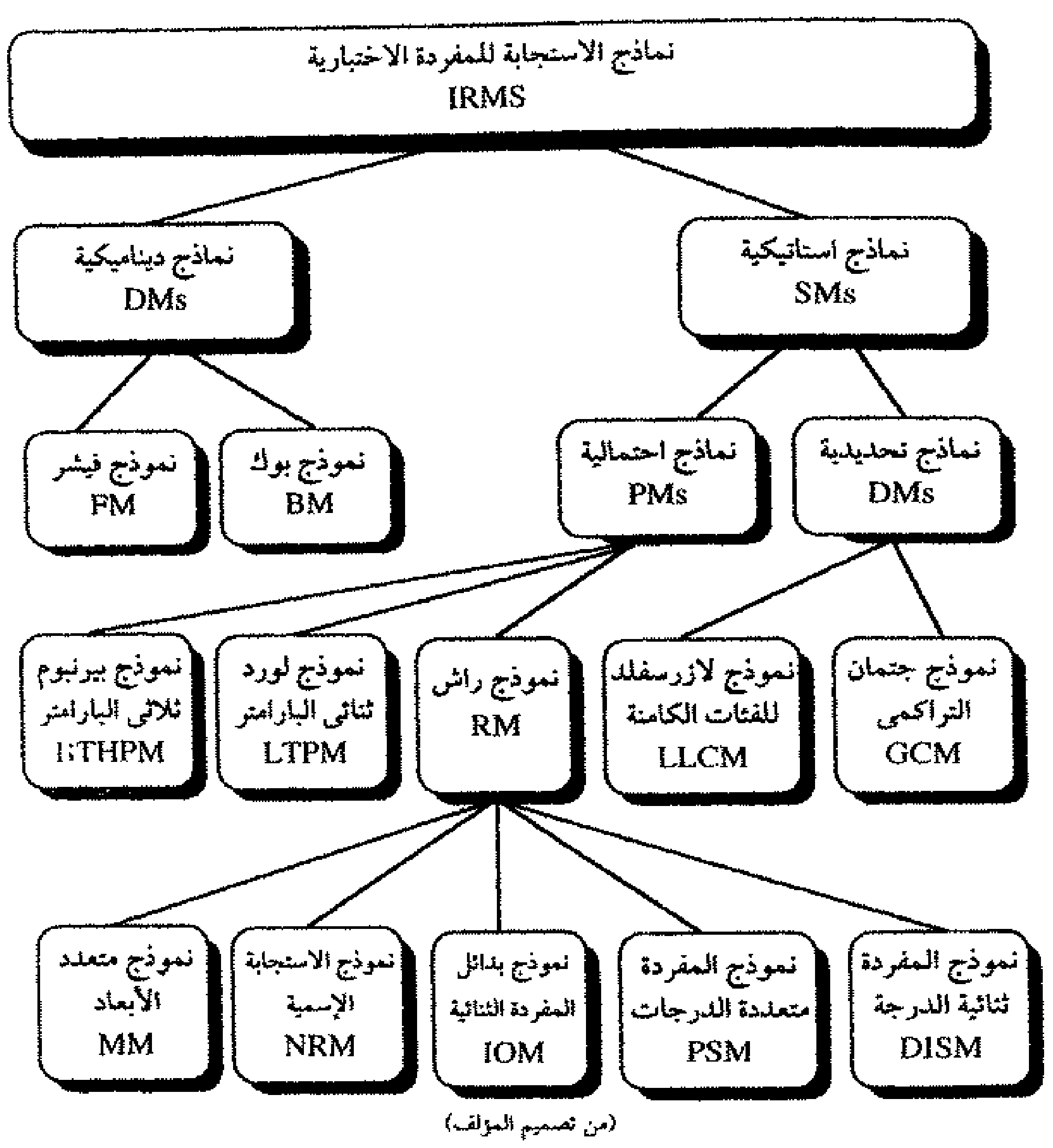
(أ) مستوى الاستجابة Response Level .

(ب) التركيب البارامترى Parametric Structure .

(جـ) الفرضيات الإحصائية Statistical Assumptions .

(د) أبعاد الفضاء الكامن Latent Space Dimensionality .

ويقترح المؤلف التصنيف التالي الذي يوضحه الشكل التخطيطي (١٤ ـ ١):



(من تصميم المؤلف)

شكل (١٤ ـ ١) يوضح تصنيف مقترح للنوعين الرئيسين لنماذج الاستجابة للمفردة الاختبارية

ويعتمد تصنيف نماذج الاستجابة للمفردة الاختبارية الموضح بشكل (١٤ ـ ١) على أربعة محكات كالتالى :

(١) **مستوى الاستجابة:** ويعد أهم المحكات التصنيفية لهذا النوع من النماذج، لأنه يتعلق بطبيعة البيانات المراد تحليلها. ويمكن التمييز بين ثلاثة مستويات للاستجابة لمفردات الاختبارات والمقاييس:

(أ) **مستوى الاستجابة الثنائية** Dichotomous Response Level: ونقصد بذلك أن تكون الاستجابة إما صحيحة أو خطأ، ويحدث ذلك عندما تكون مفردة الاختبار من نوع الاختيار من متعدد، أو الصواب أو الخطأ، أو ماشابه ذلك. غير أنه إذا كانت الاستجابة المحتملة لنوع ما من أنواع المفردات أكثر من استجابتين، فإنه يمكن اختزالها إلى استجابة ثنائية. وكثير من الاختبارات العقلية والتحصيلية تتطلب هذا المستوى من الاستجابة.

(ب) **مستوى الاستجابة المتعددة** Polytomous Response: وهنا لانستطيع اختزال الاستجابة إلى قسمين كما سبق أن أشرنا، فالاستجابة في هذا المستوى تشتمل على عدة أقسام. ومثال ذلك الاستجابة على بعض الاستبيانات، وموازين التقدير، ومقاييس الشخصية، وأسئلة المقال في اختبارات التحصيل.

(جـ) **مستوى الاستجابة المتصلة** Continuous Response Level: وفي هذا المستوى يُفترض أن عدد أقسام الاستجابة لانهائي. وهذا النمط من الاستجابات يتطلب أنواعا من مفردات الاختبار تختلف اختلافًا ملحوظًا عن تلك التي تتطلبها كل من الاستجابات الثنائية والمتعددة. ويفيد النموذج المتعلق بهذا المستوى من الاستجابة عندما يطلب من الفرد مثلا الحكم على صحة قضايا معينة حكما احتماليا ذاتيا.

(٢) **التركيب البارامتري:** ونقصد بذلك عدد وطبيعة بارامترات كل من هذه النماذج الاستاتيكية. وهذه البارامترات تتعلق بالافتراضات الخاصة بالعمليات السيكلوجية التي تنطوي عليها استجابات الفرد لمفردة اختبار.

وتوجد في الوقت الحالي ثلاثة نماذج من النوع الثنائي الذي أشرنا إليه تختلف في عدد بارامتراتها. فهناك النموذج أحادي البارامتر الذي يعرف باسم نموذج راش Rasch Model (RM)، وهو ينطبق على البيانات المستمدة من الاختبارات إذا اختلفت مفردات كل منها في الصعوبة فقط. وبذلك توصف المفردة ببارامتر واحد هو بارامتر الصعوبة. وكذلك نماذجه المنبثقة عنه الموضحة بشكل (١٤ ـ ١). وهناك النموذج ثنائي البارامتر Two - Parameters Model (LTPM) الذي ينسب إلى لورد (Lord, 1968)، وهو ينطبق على البيانات المستمدة من الاختبارات التي تختلف مفرداتها في كل من الصعوبة والتمييز، وبذلك توصف المفردة ببارامترين.



وقد أضاف بيرنبوم (Birnbaum, 1968) بارامترًا ثالثًا إلى نموذج لورد يتعلق بإمكانية التوصل إلى الإجابة الصحيحة عن مفردة اختبار عن طريق التخمين العشوائي، أي بارامتر التخمين، أو مايعرف ببارامتر الخط التقاربي الأدنى Lower Asymptote Parameter . وبذلك يكون لدينا نموذج ثلاثي البارامتر Three - Parameter Model (BTHPM) . وسوف نتناول هذه النماذج الثلاثة بمزيد من التفصيل بعد قليل.

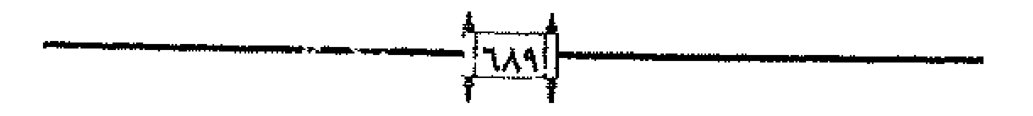
كما يوجد نموذجان من النوع المتعدد الاستجابة: أحدهما النموذج متدرج الاستجابة Graded-Response Model، والآخر النموذج نوعي الاستجابة Nominal-Response Model .

وينطبق النموذج الأول عندما تكون الاستجابات المختلفة مرتبة، ويكون لكل قسم من أقسام الاستجابة لمفردة نفس القدر من التمييز. ويمكن أن يحدث ذلك عندما تنطوى الاستجابة على عملية مرحلية أو تجمعية. وعندئذ يستخدم النموذج المتدرج الاستجابة في التنبؤ بالاستجابة المحتملة للفرد التى تتسق مع مستواه النمائى. ومثال ذلك الاستجابات على استبيانات النمو الخلقى الذى يفترض أنه يحدث من خلال مراحل متتابعة مرتبة أثناء نمو الأطفال. أما فى حالة النموذج الثانى لاتكون الاستجابات مرتبة، ولايفترض وجود تتابع مرحلى تنطوى عليه الاستجابة. أى أن ميزان الاستجابة لايكون رتبيا، بل اسميا أقسامه شاملة ومتنافية مثنى مثنى، كما أنه لايفترض فى هذه الحالة أن يكون لكل قسم من أقسام الاستجابة القدر نفسه من التمييز .

(٣) الفرضيات الإحصائية: تتعلق بطبيعة توزيع البواقى Residuals، ونقصد بذلك الأخطاء الناجمة عن عدم مطابقة البيانات لأحد هذه النماذج. فإذا كان توزيع البواقى اعتداليا Normal تُستخدم النماذج الكلاسيكية المعتادة، أما إذا كان التوزيع لوغاريتميا Logistic تستخدم نماذج الاستجابة للمفردة الاختبارية.

ومن الجدير بالذكر أن هذين النوعين من التوزيعات يمثلان أيضا الصيغتين الرياضيتين لما يُطلق عليه «المنحنى المميز للمفردة Item Characteristic Curve (ICC) الذى يُعد من المفاهيم الأساسية فى هذه النظرية المعاصرة.

(٤) أبعاد الفضاء الكامن: يمكن التمييز بين نماذج الاستجابة للمفردة الاختبارية فى ضوء أبعاد الفضاء الكامن Latent Space للاستجابة. فمعظم النماذج الموضحة بشكل (١٤ ـ ١) تفترض وجود سمة أو قدرة واحدة تفسر أداء الفرد فى الاختبار؛ لذلك تسمى نماذج أحادية البُعد Unidimensional Models، ومن أمثلتها نموذج جتمان التراكمى Gutman Cumulative Model (GCM) الذى سبق أن أوضحناه فى



الفصل الثانى عشر، وكذلك نموذج لازرسفلد للفئات الكامنة Lazarsfeld Latent Class Model (LLCM)،ويُعدان من النماذج التحديدية Deterministic Models وليست احتمالية. أما نماذج راش،والنموذج ثنائى البارامتر،والنموذج ثلاثى البارامتر فتُعد نماذج احتمالية أحادية البُعد. وهناك نماذج أخرى يُفترض أن أداء الفرد فى الاختبار يمكن أن ينطوى على سمات أو قدرات متعددة وليست سمة واحدة، لذلك تسمى نماذج متعددة الأبعاد (Multidimensional Models (MM . ويقوم خبراء القياس بإجراء دراسات وبحوث متعددة لتوسيع نموذج راش بحيث يمكن استخدامه فى تصميم الاختبارات متعددة الأبعاد وتحليل مفرداتها وبخاصة فى مجال قياس القدرات. كما توجد نماذج أخرى لم يكتمل بناؤها، كما لم تتحدد فرضياتها الإحصائية تحديدا دقيقا لأنها لاتزال موضع البحث والتجريب (Whitely, 1980) . لذلك فإن النماذج أحادية البُعد هى الشائعة الاستخدام فى القياس النفسى والتربوى فى وقتنا الحاضر.

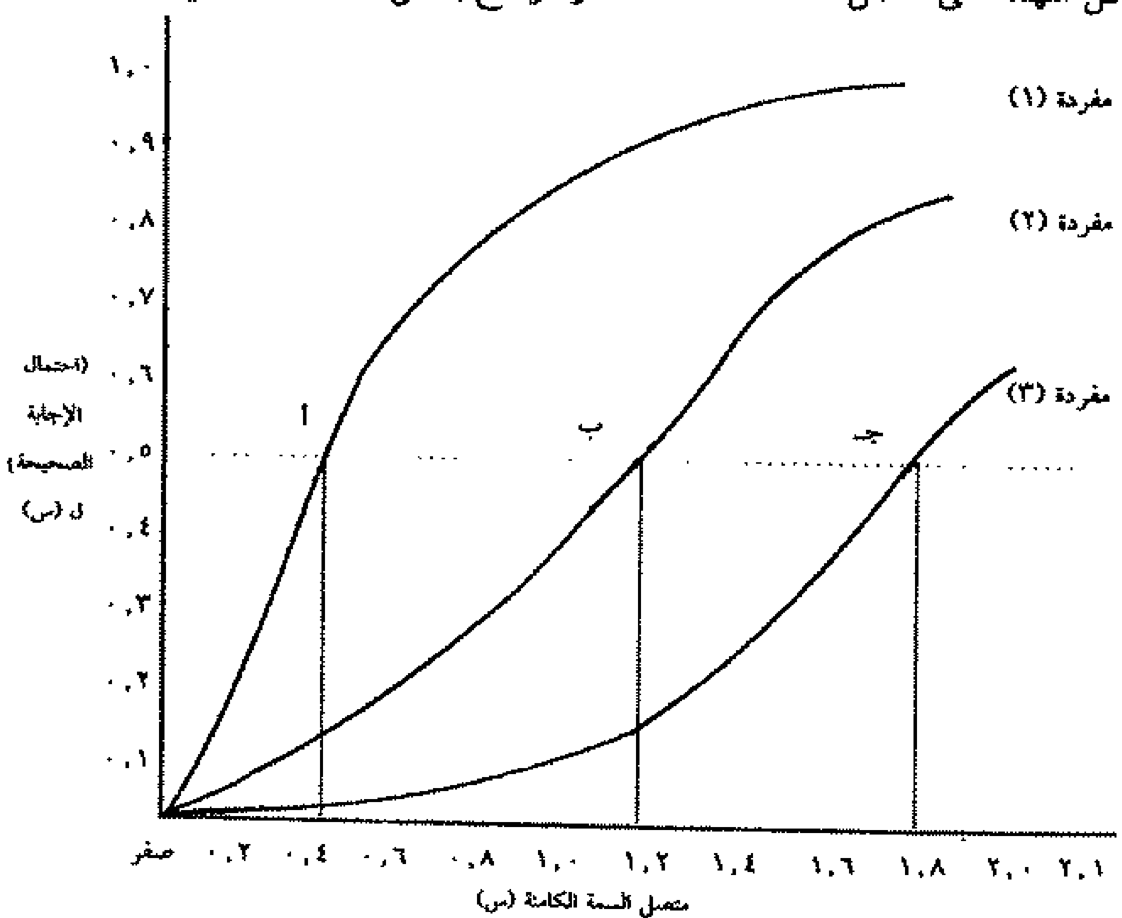
مفهوم المنحنى المميز للمفردة (ICC):

يعد مفهوم المنحنى المميز للمفردة من المفاهيم المهمة فى نماذج الاستجابة للمفردة الاختبارية التى تتناول سمة كامنة متصلة يتم قياسها بمفردات اختبارية من النوع الثنائى (أى تتطلب إجابة صحيحة أو إجابة خطأ). ويمثل هذا المنحنى بدالة رياضية تربط بين احتمال نجاح الفرد فى إجابة المفردة، وبين السمة أو القدرة التى تقيسها مجموعة مفردات أو يقيسها اختبار معين. أى أنها دالة انحدار درجة المفردة على السمة الكامنة التى يقيسها الاختبار، وهذه الدالة غير خطية Non - Linear . ويمكن التمييز بين هذه النماذج وفقا للصيغة الرياضية للمنحنى المميز للمفردة، حيث إن هذه الصيغة تمثل النموذج تمثيلا رياضيا.

ويمكن أن يتضح مفهوم المنحنى المميز للمفردة إذا أجرينا مقارنة بين كل من النظرية الكلاسيكية للاختبارات ونظرية الاستجابة للمفردة الاختبارية. فكل منهما يعتمد على فرضية مؤداها أن موقع الفرد على متصل سمة كامنة معينة يحكم احتمال إجابته إجابة صحيحة عن مفردة اختبار يقيس هذه السمة، ولكنهما يختلفان فى كيفية تحديد هذا الموقع. وارتباطه باحتمال الإجابة الصحيحة عن المفردة. أى أن الفرق بينهما يتعلق بشكل وخصائص الدالة التى تُحدد هذا القيد. فالنظرية الكلاسيكية تفترض أن منحنى الدالة المميزة للمفردة لايختلف باختلاف مفردات الاختبار، وتمثل هذه الدالة بالمنحنى الاعتدالى التراكمى لمتغير عشوائى. ويُفترض فى هذا المنحنى أن احتمال

الإجابة الصحيحة عن مفردة اختبارية دالة تزايدية مُطردة لموقع الفرد على متصل السمة الكامنة.

ونظرية الاستجابة للمفردة الاختبارية لاتتفق مع النظرية الكلاسيكية فى افتراضها تساوى احتمالات الإجابة الصحيحة عن جميع مفردات الاختبار لفرد ذى قدرة معينة. فمن الممكن أن تكون بعض المفردات أكثر صعوبة من غيرها بالنسبة لهذا الفرد؛ لذلك فإن احتمال الإجابة الصحيحة عن مجموعة من المفردات يعد دالة لموقع كل من الفرد والمفردة على متصل السمة الكامنة التى تقيسها هذه المفردات. أى أن هذه النظرية تفترض وجود دالة مميزة Characteristic Function تعتمد على موقع الفرد والمفردة معا على متصل سمة كامنة معينة، ويتحدد هذا المنحنى أو هذه الدالة إذا علمنا هذا الموقع. لذلك يطلق على هذا القياس «القياس المتحد أو المقترن Conjoint Measurement» . ولتوضيح ذلك نفترض أن لدينا ثلاث مفردات اختبارية يختلف موقع كل منهما على متصل السمة الكامنة كما هو موضح بشكل (١٤ ـ ٢) التالى:



شكل (١٤ ـ ٢) المنحنيات المميزة لثلاث مفردات وفقا لنموذج راش

ويتضح من شكل (١٤ ـ ٢) أن المفـردات الثلاث تقع عند النقط ٤ , ٠ ، ٢ , ١ , ٢ , ٢ على متصل السمة الكامنة، فالأفراد الذين يقعون عند إحدى هذه النقط على نفس المـتصـل يكون احـتمـال إجابـتهم إجـابة صـحيـحة عــن المـفردة المــناظرة لنـفس الموقع ٠,٥٠ . لذلك فإن صعوبة المفردة فى إطار نمـاذج الاستجابة للمفردة الاختبارية ليست نسبـة عدد الأفراد الذين يجيبون عنهـا إجابة صحيحة كما هـو الحال فى النظرية الكلاسيكية، وإنما هى النقطة التى على متصل السمة الكامنة التى نتوقع عندها أن يكون احتمال إجابة فرد عنها إجابة صحيحة بدون تخمين ٠,٥٠ .

فالفـرد الذى يقع عند النـقطة ٠,٤ على متـصل السمة الـكامنة يكون احـتمال إجابته إجابة صحيحة على المفردة (١) ٠,٥٠ بينما لايكون الحال كذلك عند محاولته الإجابة عن أى من المفردتين (٢) أو (٣) مثلا، وذلك لأن المفردة (١) أقل صعوبة من كل من المفردتين (٢)، (٣) بالنسبة لدرجته فى السمة الكامنة. وهذا يدل أيضا على أن المفردات الثلاث تختلف فى تمـييزها بين المستويات المختلفة للسمـة. فتمييز المفردة فى إطار نظـرية الاستجـابة للمـفردة الاختـبارية يخـتلف أيضا عـن نظيره فى الـنظرية الكلاسيكية. فلكى نحدد معامل تمييز كل من المفردات الثلاث فى شكل (٢) يجب أن نرسم خطا أفقيا عـند النقطة ٠,٥٠ على المحور الرأسى التى تناظر الإجـابة الصحيحة بحيث يقطع المنحنيات المميزة لهذه المفردات. ومعامل تمييز كل مفردة عندئذ هو ميل المنحنى المميز لها عند نقطة التقـاطع. فإذا نظرنا إلى الشكل نجد أن المفردة (١) أكثر تمييـزا من المفردة (٢)، والمـفردة (٢) أكثر تمـييزا من المفـردة (٣)، وذلك لأن ميل المنحنيات الثلاثة عند النقط أ، ب، جـ يزيد عن الواحد الصحيح بالنسبة للمفردة (١)، ويقل عن الواحد الصحيح بـالنسبة للمفردة (٢). فمعامل التمـييز المرتفع يدل على أن الفرد الذى تفوق سـمته الكامنة صعوبة المـفردة على نفس متصل السـمـة يكون احتمال إجابتـه اجابة صحيـحة عن هذه المفردة كـبيراً ، والفرد الـذى تقل سمته الـكامنة عن صعوبة المـفردة يكون احتمال إجابـته إجابة خطأ عن المفـردة كبيرا. ومعامل التـمييز المنخفـض يدل على أن احتمال الإجـابة الصحيحـة عن المفردة منتظـم، وأن المفردة لاتميز جيدا بين مستويات السمة المعينة.

ومن هذا يتضح أن لكل من النظرية الكلاسيكيـة للاختبارات، ونظرية الاستجابة للمفردة الاختـبارية افتراضات مختلفـة حول شكل وخصائص الدالة المميـزة للمفردة. فبـينما تفـترض النظريـة الكلاسيكـية أن هذه الدالة تـتخذ شكل الـمنحنى الاعـتدالى التراكمى، تفـترض نظرية الاستجابة للمـفردة الاختبارية أن الدالة تتخـذ شكل المنحنى اللوغـاريتمى التراكـمى، وأن لكل مفـردة منحنى خـاصا بها يخـتلف فى البارامـترات المحددة له، من مثل ذلك الصعوبة والتمييز.

النماذج الإستاتيكية اللوغاريتمية أحادية البُعد:

نظرا لشيوع استخدام النماذج الاستاتيكية أحادية البُعد في تصميم وبناء الاختبارات والمقاييس التربوية والنفسية في وقتنا الحاضر، فسوف نلقي الضوء على بعض أهم هذه النماذج التي أوضحناها في شكل (١٤ ـ ١). وبالطبع لن نستطيع في هذا الكتاب تناول هذه النماذج بالتفصيل أو عرض صيغها الرياضية نظرا لأن كلاً منها يحتاج إلى أكثر من مجلد. وقد ناقش المؤلف كثيرا من هذه النماذج بمزيد من التفصيل في مؤلفين آخرين، وفي عدد من الدراسات التي أجراها في هذا المجال.

نموذج راش Rasch Model:

يُطلق على هذا النموذج «نموذج البارامتر الواحد One - Parameter Model»، وقد اقترحه جورج راش G. Rasch كما أوضحنا من قبل. ويهتم بتحديد موقع المفردة الاختبارية على ميزان صعوبة جميع المفردات التي تشكّل الاختبار (بارامتر الصعوبة)، كما يهتم بتدريج مستويات قدرة الفرد باختبار معين على نفس ميزان تعيير المفردات. لذلك يعتمد النموذج على ما يسمى القياس المتحد Conjoint Mearurement ويفترض النموذج تساوي جميع مفردات الاختبار في التمييز بين مستويات القدرة المقاسة، وهي فرضية تبدو غير واقعية في تصميم الاختبارات وبنائها، كما يفترض النموذج في أبسط صورة أن تكون القدرة المقاسة أحادية البُعد، أي أن جميع مفردات الاختبار التحصيلي مثلا تقيس قدرة واحدة مثل القدرة على فهم المقروء، أو رسم الأشكال الهندسية... إلخ، وأن تكون جميع مفردات الاختبار من النوع ثنائي الدرجة، مثل مفردات الاختيار من متعدد ذات الإجابة الصحيحة الواحدة، وكذلك يفترض النموذج أن الفرد لا يلجأ إلى التخمين العشوائي في إجابته عن مفردات الاختبار.

ويوجد نموذجان أكثر عمومية ولكنهما ينتميان إلى عائلة نموذج راش في القياس التربوي والنفسي: أحدهما يسمى «نموذج التقدير المتدرج Partial Credit Model» ويُعد امتدادا لنموذج راش ثنائي الدرجة (صفر، ١). ويُستخدم هذا النموذج إذا كانت الدرجات تقع في أكثر من قسمين مُرتّبين Polytomus Scores كما في درجات أسئلة المقال وحل المشكلات، فهنا لا تكون الإجابة إما صحيحة أو خطأ وإنما تقدَّر درجات للمعلومات الجزئية المتباينة، والآخر يسمى نموذج راش متعدد الأبعاد Multidimensional Rasch Model، ويُستخدم في حالة قياس الاختبار لأكثر من بُعد واحد.

نموذج البارامترين Two-parameters Model:

ويسمح بأن تختلف مفردات الاختبار فى صعوبتها وتمييزها وهو ما يلاحظ عادة فى بناء الاختبارات. إذ من الصعب إيجاد مجموعة من المفردات تميز بدرجة واحدة بين مستويات السمة أو القدرة التى يقيسها اختبار معين، وهو الافتراض الذى اعتمد عليه نموذج راش.

نموذج البارامترات الثلاثة Three- Parameters Model:

وهو يعرف بنموذج بيرنبوم Birnbaum model، فقدأضاف بيرنبوم بارامتراً ثالثاً وأطلق عليه بارامتر الخط التقاربى الأدنىLower Asymptote Line أو بارامتر التخمين Guessing Parameter فيما يتعلق بالمفردات الاختبارية التى تتطلب الاختيار بين بدائل معطاة. وهذا البارامتر يحدد احتمال أن يجيب فرد يفترض أن مستوى قدرته منخفضا انخفاضًا لا نهائيًا، ومع هذا يجيب إجابة صحيحة عن غالبية مفردات الاختبار عن طريق التخمين. لذلك تبدو أهمية هذا البارامتر عند مطابقة بيانات مستمدة من مجموعات من مفردات الاختبارية من نوع الاختيار من متعدد لهذا النموذج .

ومن الجدير بالذكر أن هناك نماذج أخرى انبثقت عن هذه النماذج الثلاثة تتعلق بتحليل مفردات موازين التقدير Rating Scales، ونموذج الاستجابة الاسمية Nominal Response Model، ونموذج البدائل التى تشتمل عليها مفردات الاختيار من متعدد Item Options Model، والنموذج رباعى البارامتر Four-parameters Model.

ولكل من النماذج السابقة صيغ رياضية لامجال لذكرها هنا، وهذه الصيغ عبارة عن دوال رياضية لوغاريتمية احتمالية تربط بين بارامترات النموذج. وتُستخدم أساليب التحليل العددى Numerical Analysis فى تقدير بارامترات كل نموذج، مثل بارامتر صعوبة المفردة، وبارامتر تمييز المفردة، وبارامتر قدرة الفرد، وبارامتر التخمين، وكذلك التحقق من حسن مطابقة البيانات للنموذج باستخدام بعض الأساليب الإحصائية. ويتطلب تنفيذ هذه الأساليب برامج خاصة للحاسوب. ويوجد فى الوقت الحاضر عدد من برامج الحاسوب المتوافرة التى يمكن باستخدامها التوصل إلى قيم تقديرية لبارامترات النموذج الذى يرى الباحث ملاءمته لاختبار أو مقياس معين. ومن أمثلة هذه البرامج الشائعة الاستخدام مايلى:

(١) برنامج RASCAL (WINDOWS) : يُجرى تحليل مفردات الاختبارات باستخدام نموذج راش، ويُعد برنامجا قويا وسهل الاستخدام ويعمل على بيئة النوافذ. ويمكن استخدامه في تعيير مفردات اختبارية يبلغ عددها ٧٥٠ مفردة بدون حد أقصى لعدد الأفراد الذين يُختبرون بهذه المفردات.

(٢) برنامج RSP (1.0) : يُجرى تحليل المفردات باستخدام نموذج راش أيضا، ولكن في التصميمات غير الكاملة التي تشتمل على عدد من الاختبارات يتكون كل منها من عدد من المفردات تُطبق على مجموعة من الأفراد. ويجب أن تلتحم الاختبارات معًا بمفردات مشتركة أو بتوزيعات مشتركة للسمات المراد قياسها. ويستوعب البرنامج ٣٢ اختبارًا ، ويتكون كل منها من ٩٦ مفردة على الأكثر.

(٣) برنامج XCALIBRE (WINDOWS) : يُجرى العمليات اللازمة لتقدير بارامترات النموذجين ثنائي وثلاثي البارامتر. ويستوعب ٧٥٠ مفردة اختبارية بدون حد أقصى لعدد الأفراد الذين يُختبرون بهذه المفردات.

(٤) برنامج ASCAL : يُجرى أيضا تقديرات بارامترات النموذجين ثنائي وثلاثي البارامترا ولكن باستخدام إحصاء بييز Bayesian Approach ،ويستوعب ٢٥٠ مفردة اختبارية يختبر بها ١٥٠٠٠ فرد على الأكثر.

(٥) برنامج MSP 3.0 : يُجرى تحليل مفردات متعددة الاستجابات Polytomous Items وكذلك المفردات ثنائية الاستجابة Dichotomous Items، وذلك لتصميم اختبارات أحادية البُعد تنطبق البيانات المستمدة منها على أحد نماذج الاستجابة للمفردة الاختبارية. ويستوعب ١٠٠ مفردة اختبارية على الأكثر يُختبر بها عدد لايزيد عن ٣٢٠٠٠ فرد، ويجرى تقديرات بارامترات النموذج باستخدام أساليب لابارامترية.

(٦) برنامج PARELLA 1.1 : يُجرى تحليل لاستجابات الأفراد لمجموعة من فقرات أو عبارات موازين الاتجاهات والتفضيلات، حيث يتم تمثيل موقع كل فرد بالنسبة للفقرة التي يوافق عليها بأقصر مسافة بينهما في بُعد واحد. ويستوعب ٦٠ فقرة يستجيب لها ٣٠٠٠ فرد على الأكثر.

مزايا وعيوب النماذج الإستاتيكية:

مما سبق يتبين تعدد النماذج الإحصائية السيكومترية التي يمكن استخدامها في القياس النفسي والتربوي وتنوعها وتباينها. ونظرا إلى أن النماذج الإستاتيكية المتعلقة بالاستجابة للمفرد الاختبارية (SIRMs) نالت اهتمامًا كبيرًا من المشتغلين ببناء



الاختبارات والمقاييس، فإنه يجب على المهتمين بذلك أن يتعرفوا مزايا كل من هذه النماذج وعيوبها، وأن يختاروا النموذج الذى يتميز بالبساطة والواقعية وإمكانية التنفيذ، ومدى تحقق فرضياته فى البيانات، وسهولة تقدير بارامتراته، ومنطقية تفسير نتائجه.

وعلى الرغم من أن نموذج راش هو أكثر هذه النماذج استخداما فى بناء الاختبارات والمقاييس لما يتوافر فيه من بعض هذه المزايا، إلا أنه يجب على من يستخدم هذا النموذج فى تعيير المفردات الاختبارية أن يكون مدركا لجوانب القصور فيه وحدود تطبيقاته، حتى لا يُساء استخدامه أو تُعمم نتائجه تعميماً خطأ.

ولعل أهم مزايا وعيوب هذا النموذج ما يلي:

(١) نموذج راش هو النموذج الوحيد من بين النماذج الاحتمالية الذى تتوافر فيه خاصية استقلالية بارامتراته، أى عدم تباين Invariance تقديرات قدرة الأفراد بتباين عينة المفرادت التي يُختبرون بها، وعدم تباين الخصائص السيكومترية للمفردات بتباين عينة الأفراد التى أُجرى عليها الاختبار. غير أن هذا يتطلب التحقق بصورة دورية من عدم حدوث تغيرات ربما تطرأ على القيم التقديرية لبارامترات النموذج نتيجة لتغير خصائص أفراد مجتمع المختبرين، أو تغير طبيعة أو محتوى أو معنى المفردات الاختبارية بالنسبة إليهم بمرور الزمن.

(٢) يعد نموذج راش أبسط النماذج الإستاتيكية الاحتمالية التى أشرنا إليها، حيث يشتمل على بارامترواحد يتعلق بمفردات الاختبار وهو صعوبة المفردة Item Difficuiry، فعلى الرغم من تباين صعوبة مفردات الاختبار فإن النموذج يفترض أن جميع المفردات تتساوى فى قدرتها على التمييز. وهذا القيد يُضيف قيداً آخر على انتقاء المفردات التى يشتمل عليها الاختبار. فالمفردات لا يجب أن تميز بين مستويات القدرة المقاسة تمييزا دالا فحسب، بل إن جميع مفردات الاختبار يجب أن تكون قدرتها على التمييز متساوية تقريبا.

ويرى مؤيدو هذا النموذج أن الميزة التى تنجم عن ذلك هى أن المجموع غير الموزون لدرجات مفردات الاختبار يعد إحصاءً كافيًا Sufficient Statistic لتقدير قدرة الفرد المختبر. كما أنه يترتب على ذلك أيضا ميزة أخرى تتعلق بوصْل مفردات الاختبارات على شكل حلقات متتابعة Item Linking لتحقيق عملية تكافؤ الاختبارات Test Equating وهو أمر ضرورى بل وأساسى فى بناء نظم بنوك الأسئلة من أجل إمكانية تشكيل العديد من صيغ الاختبارات المختلفة، ولكن تؤدى إلى درجات كلية على ميزان موحد. فعملية التوصل إلى درجات متناظرة للصيغ المختلفة من الاختبار



تسمى عملية التكافؤ. وعلى الرغم من أن تكافؤ الدرجات الحقيقية سواء التكافؤ الأفقى أو الرأسى يمكن التوصل إليه بشروط متعددة، إلا أن نماذج الاستجابة للمفردة الاختبارية وبخاصة نموذج راش ربما ييسر ذلك. غير أن هذا يتطلب التحقق من فرضية تساوى المفردات فى قدرتها على التميز، وذلك بفحص أنماط البواقى المعيارية التى تنجم عن تقييم حسن مطابقة البيانات للنموذج.

(٣) يؤدى نموذج راش إلى منحنيات مميزة للمفردات (ICCs) تكون متوازية أى لا تتقاطع، وكذلك المنحنيات المميزة للأفراد المختبرين (PCCs). وهذا ييسر عملية التقدير الإحصائى لبارامترات النموذج **Parameters Estimation** والوصول إلى ما يسمى بالتقارب **Convergence**. فالمنحنى المميز للمفردة هو دالة رياضية تربط بين احتمال الإجابة الصحيحة لفرد ما عن مفردة اختبارية، وبين القدرة التى يقيسها الاختبار. والمنحنى المميز للفرد هو تلك الدالة التى تربط بين احتمال الإجابة الصحيحة لمجموعة من الأفراد المتساوين تقريبا فى القدرة المقاسة وبين صعوبات المفردات.

غير أن توازى هذه المنحنيات لا يحدث إلا إذا حققت البيانات الفرضيات التى يستند إليها النموذج، وبخاصة فرضية البعد الأحادى للاختبار، وتساوى تمييز مفرداته، وعندئذ يمكن تفسير مستويات وأنماط القدرة المقاسة. علما بأن التحقق من خاصية البُعد الأحادى يشوبه كثير من المشكلات الإحصائية والسيكومترية، كما أن كثيرا من الاختبارات ربما لا تقيس سمات أحادية البعد.

(٤) فى حين تتميز النماذج الاستاتيكية الأخرى وبخاصة النموذج ثلاثى البارامتر بمزايا تفوق نموذج راش، حيث إنه يحدد بارامتر الموضع **Location Parameter**، وبارامتر ميل المنحنى المميز **Slope Parmeter**،وبارامتر التخمين **Guessing** لكل مفردة وبخاصة مفردات الاختيار من متعدد التى يعتمد عليها نموذج راش مما يؤدى إلى نتائج أكثر واقعية فى العملية الاختبارية وبناء المقاييس، إلا أن هذا يؤدى إلى تقاطع المنحنيات المميزة للمفردات، وبالتالى يؤدى إلى صعوبات فى عمليات التقدير الإحصائى **Parameters Estimation** لهذه البارامترات. لذلك نجد أن المنحنيات المميزة فى حالة استخدام نموذج راش تتابها بعض الذبذبات والتعرجات بالنسبة إلى الأفراد منخفضى القدرة الذين يكون احتمال توصلهم للإجابة الصحيحة عن مفردة ما من مفردات الاختيار من متعدد التى يشتمل عليها الاختبار أقل مما هو متوقع عن طريق الصدفة، نظرا إلى أنهم ينجذبون إلى المشتتات غير الصحيحة للمفردة، وذلك

لعدم أخذ الأخطاء الناجمة عن التخمين بعين الاعتبار في النموذج، وبخاصة أنه يعتمد اعتمادا كليا على مفردات الاختيار من متعدد. لذلك اتجهت البحوث في الآونة الأخيرة إلى بناء نماذج بديلة لنموذج راش تعالج هذه الظاهرة. وهذا أمر ضروري نظرا إلى أن مشتتات مفردات الاختيار من متعدد Distractors تبنى من أجل الكشف عن أمثال هؤلاء الأفراد.

أما إذا اشتمل الاختبار على أسئلة من النوع الذي يتطلب إجابة حرة فإنه يجب اختيار نموذج آخر مناسب لهذا الغرض، مثل نموذج التقدير المتدرج Partial Credit Model.

غير أن كثيرا من خبراء القياس المعاصرين يرون أن هذه النماذج سوف تؤدي إلى زيادة المرونة والدقة في بناء الاختبارات التحصيلية. فمثلا يمكن تجميع عدد كبير من المفردات الاختبارية التي يتم تعييرها Calibrated Items وفق أحد هذه النماذج، مما يجعل الخصائص الإحصائية لهذه المفردات مستقلة عن عينة الطلاب التي تم تجريب المفردات عليها. وهذا بالتالي يجعل من الممكن بناء اختبارين يؤديان إلى درجات متكافئة دون الحاجة إلى تطبيقهما معا، أو تصميم اختبار ذي خصائص سيكومترية محددة مسبقا، أو تحديد ما إذا كان أحد الطلاب قد حقق مستوى محدد من مستويات الإتقان، وهذا المستوى يكون مستقلا عن مجموعة المفردات التي أجاب عنها الطالب. أي إذا اختبر الطالب باختبار آخر يشتمل على عينة من المفردات المسحوبة من نطاق هذه المفردات Item Domain والتي تم تعييرها وفق أحد هذه النماذج، فإنه يتوقع أن يحقق الطالب مستوى الإتقان السابق نفسه تقريبا في حدود قدر ضئيل من الخطأ.

ومما لا شك فيه أن تقنيات الحاسوب أثرت استخدم نماذج الاستجابة للمفردة الاختبارية وأدت إلى مزيد من تطبيقاتها في مختلف المجالات، مثل بناء الاختبارات مرجعية المحك، وتفريد القياس، وبناء نظم بنوك الأسئلة، والتكافؤ الأفقي والرأسي لدرجات الاختبارات، وغير ذلك. وسوف نلقي الضوء على بعض هذه التطبيقات في الفصل التالي.

(ثانيا) نظرية الثبات متعدد الأوجه Generalizability Theory:

أوضحنا في الجزء السابق بعض جوانب قصور النظرية السيكومترية الكلاسيكية للاختبارات (CTT)، وكيف حاولت نظرية الاستجابة للمفردة الاختبارية (IRT) التغلب على ذلك. وعلى الرغم من أن كلا من النظريتين تعالج تباين أخطاء القياس معالجة

مختلفة، إلا أن كليهما ينظر إلى مصدر هذه الأخطاء نظرة ضيقة ومحدودة. فالمعلومات أو البيانات المستمدة من أى إجراءات قياس سواء فى العلوم الطبيعية أو العلوم السلوكية تكون مشوبة بدرجة معينة من الخطأ. فقياس الطول مثلا يختلف وفقا لظروف متعددة متعلقة بهذا القياس، مثل المسطرة المستخدمة والمادة المصنوعة منها، والفرد القائم بالقياس، وتسجيل القيمة العددية المناظرة للطول، وظروف الإضاءة، وغير ذلك. ولزيادة دقة القياس يمكن إيجاد متوسط قياسات متعددة للطول فى ظروف محددة للتوصل إلى قيمة تقديرية للقياس المثالى، أى القياس الافتراضى الذى يمكن الحصول عليه إذا استطعنا إيجاد المتوسط فى جميع الظروف المحددة الممكنة. غير أن هذا يتطلب معرفة عدد مرات القياس والظروف المختلفة المتعلقة به للتوصل إلى درجة مقبولة من الدقة. فإذا رأينا أن اختيار مسطرة مصنوعة من مادة معينة لايؤثر على قياس الطول، ولكن ربما يؤدى اختلاف القائمين بالقياس إلى تباين ملحوظ فى القيم العددية المناظرة للطول، فإنه ربما يفضل فى هذه الحالة إيجاد متوسط القياسات باستخدام عدد قليل من المساطر، وزيادة عدد القائمين بالقياس. وعلى الرغم من أنه يمكننا استخدام مسطرة واحدة فى جميع القياسات، وبذلك تقل الأخطاء وتزيد دقة القياس، إلا أن هذا يضع قيدا على نطاق ظروف القياس الذى يمكن التعميم عليها، مما يجعل تفسير القياسات الناتجة محددا، أى أن أخطاء القياس من هذا المنظور ليست متضمنة فى عملية القياس، وإنما تعتمد على تحديد أى ظروف القياس التى تسهم فى الخطأ، لكى يتسنى تعميم القياس عليها. ولعل هذا هو ماتتناوله نظرية الثبات متعدد الأوجه.

بعض أوجه قصور المفهوم الكلاسيكى للثبات والصدق:

يعد كل من مفهوم الثبات والصدق من المفاهيم الأساسية فى القياس التربوى والنفسى وبخاصة الاختبارات والمقاييس. فالبيانات المستمدة من هذه الأدوات يجب أن تتميز بالثبات والصدق مهما اختلف الغرض من استخدامها. فالثبات من وجهة النظر الكلاسيكية يعنى أن تعكس هذه البيانات الجوانب الحقيقية للسمة أو القدرة المقاسة، أى المصادر المنتظمة للتباين، ولاتعكس عوامل الصدفة أو العشوائية، أى تباين الخطأ.

أما الصدق فيعنى مدى الاستفادة من المقاييس فى صنع قرارات تتعلق بأهداف معينة، أو مدى ملاءمة الاستدلالات من درجات الاختبارات أو غيرها من أدوات القياس. فبقدر ماتكون المعلومات التى تمدنا بها هذه الأدوات مضللة أو غير كافية لصنع قرارات معينة، تكون البيانات المستمدة منها غير صادقة مهما كانت البيانات ثابتة أو متسقة. فأدوات القياس تكون قليلة الجدوى إذا لم تكن صادقة لغرض معين.

والنظرية الكلاسيكية للاختبارات تعد من النظريات الواهنة والمحدودة Weak True Score Theory في معالجتها لمفهومي الثبات والصدق نظرا لانها لاتميز بين أخطاء القياس، وإنما تعطى قيمة تقديرية كلية لمصادر أخطاء القياس المتعددة المتعلقة بإحدى صيغ الاختبار المستخدمة بواسطة مختبر أو فاحص معين، وفي ظروف معينة؛ لذلك تتعدد أساليب تقدير الثبات كما أوضحنا في الفصل الرابع.

وقد اهتم كرونباك Cronbach ومعاونوه بمراجعة وتعميق تعريف وتفسير مفهومي الثبات والصدق ومعالجتهما معالجة أكثر شمولية من منظور متعدد الأبعاد. فقد أعاد كرونباك تفسير نظرية الثبات، وقام بوضع أسس بناء نظرية جديدة تهتم بإمكانية تعميم عينة القياسات المستمدة من مجموعة من الملاحظات على النطاق الشامل لهذه الملاحظات Universe of Observations،وأطلق على هذه النظرية «نظرية إمكانية التعميم Generalizability Theory . وقد أشرنا إلى هذه النظرية في الفصلين الرابع والخامس.

ويرى كرونباك أن ضعف النظرية الكلاسيكية للاختبارات يرجع إلى تبسيطها الشديد لمفهومي الثبات والصدق، واستخدام الباحثين لهذين المفهومين استخداما آليا، مما أدى إلى غموض معناهما الحقيقي واستخدامهما الفعلي ، ويؤكد أن معظم جوانب النظرية الكلاسيكية تعد حالات خاصة من نظريته العامة الجديدة.

فنظرية إمكانية التعميم تزودنا بإطار فكرى مرجعي لكيفية تقدير ثبات القياسات السلوكية التي نحصل عليها بواسطة إجراءات أو أدوات القياس المختلفة، مثل الاختبارات، وقوائم الملاحظة أو المراجعة، وموازين التقدير، ومقاييس الميول والاتجاهات والشخصية التى تُستخدم في البحوث التربوية والنفسية وفي اتخاذ القرارات في المجالات التطبيقية.

ونظرية إمكانية التعميم لاتعد امتداداً للنظرية الكلاسيكية فحسب، بل هى أسلوب جديد يساعد الباحث ومصمم المقياس في معرفة ما إذا كانت قياسات معينة تحقق الهدف الذى ينشده من القياس، مثل تقويم مقرر دراسى، أو انتقاء الأفراد، أو تقويم بنائى أو تجمعى لبرنامج معين، أو الحصول على أدلة تتعلق بصدق التكوين الفرضى بعامة.

بعض المفاهيم الأساسية لنظرية إمكانية التعميم:

عرض كرونباك ومعاونوه (Cronbach, Gleser, Nanda & Rajaratnam, 1972) هذه النظرية بالتفصيل في مؤلف واحد بعنوان «اعتمادية القياسات السلوكية

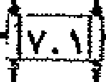
The Dependability of Behavioral Measurements : Theory of Generaizability for Scores and Profiles ».

وقد اهتم كثير من خبراء القياس بعد ذلك بتطوير النماذج الإحصائية متعددة المتغيرات التي تستند إليها هذه النظرية (Brennan, 1980; Cardinet, 1981, webb; 1981). إذ إن هذه النظرية تهتم بتحليل المكونات المتعددة للتباين في مجموعة من القياسات أو الملاحظات. وقد لاحظنا في الفصل الرابع أن بعض أساليب تقدير الثبات في إطار النظرية الكلاسيكية تستند إلى طرق تحليل التباين Analysis of Variance غير أن هذه الطرق لا تستخدم في إطار نظرية إمكانية التعميم كما هي مستخدمة في التصميمات التجريبية، حيث إنها تختلف عنها في المصطلحات وأنواع التصميمات، كما تختلف عنها في المنظور والتوجه. ويمكن استخدام هذه النظرية في مناقشة قضايا سيكومترية متعددة بطريقة منظمة ربما أغفلتها النظريات السيكومترية الأخرى أو عالجتها معالجة سطحية.

غير أن ما تتطلبه هذه النظرية من تصورات فكرية وأساليب إحصائية معقدة .بما لا يكون عائقا إذا أردنا الإفادة من إسهاماتها المتعددة.

فقد أوضحنا في الفصل الرابع أن النظرية الكلاسيكية في تفسيرها للثبات تعتمد على المقاييس المتوازية أو المتكافئة Parallel Measures. فهي تفترض وجود مقياسين متكافئين أو أكثر لقياس سمة أو قدرة معينة. ونقصد بتكافؤ المقاييس تماثل محتواها، وتساوي متوسطات وتباين درجاتها، وكذلك تساوي الارتباطات البينية لدرجات مفرداتها المتناظرة.

ويتضح استخدام هذا المفهوم في بناء الاختبارات المقننة التي تناولناها في الفصل الثامن وغيره من الفصول بدرجة أفضل من استخدامه في الأنواع الأخرى من المقاييس التي تعتمد على التقديرات Ratings كما في قياس الاتجاهات والشخصية. فعلى الرغم من أن مثل هذه الأنواع تكون متوسطات التقديرات، وانحرافاتها المعيارية، والارتباطات البينية غير متساوية، إلا أنه يجب التأكد من ثبات هذه التقديرات بطريقة ما؛ لذلك كان من الضروري ابتكار مفاهيم أخرى لا تعتمد على هذه النظرية



الكلاسيكية. ومن بين هذه المفاهيم، مفهوم النطاق الشامل Universe Domain. فنحن نهتم في التقديرات، وفي القياس بعامة، بإمكانية تعميم هذه التقديرات أو القياسات من موقف إلى آخر، ومن عينة من الأفراد إلى عينة أخرى. فالملاحظات (ولتكن درجات مجموعة من الأفراد في اختبار أو مقياس ما) التي يجمعها الباحث في موقف معين، تعد بمثابة عينة من «الملاحظات» التي يشتمل عليها نطاق شامل يتضمن «ظروف أو شروط جمع الملاحظات Conditions of Observations»، أي المثيرات أو المواقف المعينة التي يتم في ضوئها جمع الملاحظات أو الحصول على القياسات. وبالطبع يمكن أن تنتمي ملاحظات أو قياسات معينة إلى نطاقات شاملة مختلفة. لذلك إذا أراد الباحث أن يتأكد من دقة تعميم الملاحظات أو القياسات على نطاق شامل معين، فعليه أن يحدد هذا النطاق تحديدا دقيقا.

فالدرجة الحقيقية في إطار النظرية الكلاسيكية هي متوسط درجات الفرد التي يحصل عليها إذا طبق عليه عدد لا نهائي من الاختبارات المتوازية، أي أن الدرجة الحقيقية هنا تعتمد على نطاق شامل واحد للقياس. أما الدرجة الشاملة Universe score التي يقترحها كرونباك في نظريته، فتعتمد على فكرة تعدد النطاقات الشاملة التي يود صانع القرار تعميم قياساته عليها.

وهذا يبين أنه ليس هناك درجة حقيقية واحدة للفرد في الاختبار True score، كما هو معروف في النظرية السيكومترية الكلاسيكية، وإنما يكون له درجة شاملة Universe score، يمكن تعريفها بأنها القيمة المتوقعة لدرجاته الملاحظة Observed Scores التي يحصل عليها في مختلف المواقف التي تنتمي إلى النطاق الشامل المطلوب. فالدرجات التي تنتمي إلى نطاق شامل معين ربما تتباين في أكثر من جانب يطلق على كل منها «وجه أو بعد Facet». فمثلا ربما تكون الأوجه أو الأبعاد هي حجم عينة الأفراد الذين يطبق عليهم اختبار أو مقياس معين، أو أنواع التدريبات التي يتلقونها قبل اختبارهم، أو صيغة الاختبار التي تجرى عليهم، أو ظروف إجراء الاختبار، وغير ذلك. وتعد كل هذه الأوجه أو الأبعاد مصدرا هاما من مصادر التباين التي يجب أن تؤخذ بعين الاعتبار. لذلك ينبغي أن يحدد الباحث مصدر أو مصادر التباين التي يهتم بها. فمثلا ربما يميز الباحث بين عينات من الأفراد تشتمل على أعداد تتراوح بين ١، ١٥، ٣٠ فردا، وأكثر من ٣٠ فردا. أو ربما يكون الأفراد الذين يُجرى عليهم الاختبار طلابا أو معلمين أو فئات أخرى، وهكذا.

لذلك يوصي كرونباك Cronbach القائمين ببناء الاختبارات أو نشرها تحديد «الشروط أو الظروف Conditions» التي تُمكّن مستخدمي الاختبار من الحصول على نتائج متكافئة من الصيغتين (أ)، (ب) مثلا من صيغ اختبار معين. إذ ربما يكون أحد هذه الشروط أن تطبق صيغتا الاختبار على عينة تشتمل على عدد لا يقل عن ٣٠ فردا لكي يحصل على نتائج متكافئة، أو ربما يجب عليه أن يُطبق الاختبار على أفراد تلقوا قدرًا معينًا من التدريب، وما إلى ذلك.

والحقيقة أن نظرية إمكانية التعميم لم تهتم فقط بكيفية بناء أدوات قياس تكون نتائجها متكافئة في ظروف ومواقف اختبارية مختلفة، وإنما اهتمت أيضا باستخدامات أدوات القياس في صنع قرارات معينة. لذلك تميز بين نوعين من الدراسات التي تتضمنها: النوع الأول هو دراسات إمكانية التعميم Generalizability Studies ويُرمز لها بالحرف الأول (G)، والنوع الثاني هو دراسات القرارات Decision Studies ويُرمز لها بالحرف الأول (D). وهذا التمييز ليس واضحا في النظرية الكلاسيكية للاختبارات.

دراسات إمكانية التعميم (G) Generalizability Studies:

تعتبر النظرية الكلاسيكية للاختبارات أن درجة الفرد في اختبار أو مقياس ما تتكون من جزأين، هما الدرجة الحقيقية وتُمثّل الأداء أو القدرة الحقيقية، ودرجة وحيدة للخطأ وهي تعكس عوامل الصدفة. والدرجة الحقيقية للفرد هي القيمة المتوقعة لدرجاته الملاحظة في جميع المقاييس المتوازية. ونظرا لأن يُفترض أن الدرجات الحقيقية تكون مستقلة عن بعضها البعض، فإن تباين الدرجات الملاحظة يساوي مجموع تباين الدرجات الحقيقية مضافا إليه تباين الخطأ. ويُعرف الثبات في إطار هذه النظرية، بأنه مربع معامل الارتباط بين الدرجات الملاحظة والدرجات الحقيقية، وهذا يساوي النسبة بين التباين الحقيقي والتباين الملاحظ. ونظرًا لأننا لا نَعرف الدرجة الحقيقية للفرد، فبالتالي لا يمكن معرفة تأثير عوامل الصدفة في تحديد الدرجة الملاحظة. ولكن يمكن تقدير تأثير هذه العوامل على القياس بعامة. وتوجد عدة طرق لتقدير معامل الثبات بهذا المعنى التقليدي كما سبق أن ذكرنا إحداها إيجاد معامل الارتباط بين درجات الأفراد في اختبارين متوازيين، أي صيغتين متكافئتين للاختبار يُطبقان عليهم في الوقت، أو استخدام طريقة كيودر - ريتشاردسون (الصيغة ٢٠)، أو معامل ألفا لكرونباك Cronbach's Alpha Coefficient.

وهاتان الطريقتان تتطلبان تطبيق الاختبار نفسه مرة واحدة. كما توجد طرق أخرى تُستخدم في الحصول على أنواع مختلفة من معاملات الثبات، مثل معاملات

الاستقرار Stability Coefficients، ومعاملات الاستقرار والتكافؤ Stability and Equivalence Coefficients، وثبات تقديرات المحكمين Reliability of Ratings. ولكل نوع من هذه الأنواع من المعاملات تباين خاص بخطأ أحادى البعد Unidimensioal Error Variance. ونقصد بذلك أنه يوجد مصدر واحد أو تباين واحد غير متميز للخطأ خاص بكل منها. فكل طريقة من هذه الطرق تفترض فروضا مختلفة عن مكونات مصدر تباين الخطأ فى القياس.

وهذا فى الحقيقة يعد نقطة ضعف أخرى فى النظرية الكلاسيكية للاختبارات. فاختلاف طرق تقدير الثبات يؤدى إلى غموض تفسير نتائج الاختبارات والمقاييس. ويرجع سبب ذلك كما ذكرنا إلى البساطة الشديدة التى تتميز بها النظرية الكلاسيكية. وهذا هو ما دعا كرونباك وزملاءه إلى بناء نموذج متعدد الأوجه Multifacet Approach يتصف بالعمومية لمواجهة نقط الضعف السابقة فى النظرية.

فنظرية إمكانية التعميم تأخذ بعين الاعتبار المصادر المتعددة لأخطاء القياس، وتهتم دراسات إمكانية التعميم Generalizability Studies (G) بتحديد درجة تكافؤ نتائج الاختبارات التى نحصل عليها فى ظروف اختبارية مختلفة وتحت شروط متباينة، أى تهتم بدراسة تأثير معاينات هذه الشروط أو الظروف على الدرجات من أوجه مختلفة Facets. ويمكن إجراء ذلك عن طريق جمع البيانات من الأفراد المختبرين فى ظروف محددة (أى مستويات مختلفة لأوجه محددة)، وتقدير مكونات التباين التى ترجع إلى اختلاف مستويات الأوجه، وكذلك تقدير تفاعل هذه المكونات باستخدام طرق تحليل التباين Analysis of Variance، وبذلك يمكننا الحصول على معاملات تسمى معاملات إمكانية التعميم Coefficients of Generalizability. ومعامل إمكانية التعميم هو النسبة بين تباين الدرجة الشاملة Universe Score إلى تباين الدرجة الملاحظة Observed Score. أى أنه يمكننا اعتبار معامل إمكانية التعميم بديلا عن معامل الثبات الذى نحصل عليه باستخدام النظرية الكلاسيكية للاختبارات. وكل اختبار أو مقياس يكون له أكثر من معامل إمكانية تعميم واحد على حسب الأوجه التى يتم أخذها بعين الاعتبار فى دراسة إمكانية التعميم G - study. ولذلك فإن تحليل التباين فى هذه الدراسات يحتاج إلى تصميمات تجريبية Experimental Designs أكثر تعقيدا من التصميمات المستخدمة فى التحليل الكلاسيكى لبيانات الاختبارات.

دراسات القرارات Decision (D):

من المعلوم أن أدوات وأساليب القياس تُستخدم في جمع بيانات تفيد في صنع القرارات، أو في التوصل إلى استنتاجات معينة. ومثال ذلك قبول الطلاب في برامج تعليمية مختلفة، أو انتقاء الأفراد، أو التعرف على الأطفال المتفوقين في التحصيل الدراسي وغير ذلك. وقد تبين لنا مما سبق أن المعلومات التي تمدنا بها دراسات إمكانية التعميم تستخدم في جمع بيانات تفيد في تقدير مكونات تباين القياسات التي نحصل عليها بطريقة ما من الطرق، وهذا يفيد في تطوير هذه القياسات. أما دراسات القرارات فتستخدم في جمع بيانات تفيد في صنع العديد من القرارت، والتوصل إلى كثير من التفسيرات والاستنتاجات بناء على النتائج التي تتوصل إليها دراسات إمكانية التعميم G - Studies. ويمكن أن نميز بين أربعة أنواع من القرارت على الأقل هي:

(١) القرارت المطلقة:

عندما نود تصنيف أداء أحد الأفراد بطريقة معينة ينبغي تحديد مستويات الأداء التي يتم تصنيفه على أساسها. وقد عرضنا لهذا النوع من القرارت في الفصل الثامن عند مناقشتنا للاختبارات مرجعية المحك. فنحن نهتم في هذا النوع من القياس بموازنة أداء الفرد بمستوى أداء مطلق محدد مسبقا، مثل النسبة المئوية لعدد مفردات الاختبار التي ينبغي أن يجيب عنها الفرد إجابة صحيحة، أو عدد الأخطاء، أو الزمن اللازم لأداء عمل معين، وما شابه ذلك. وهنا يكون القرار خاصا بفرد معين دون الاهتمام بغيره من الأفراد. ويفيد مثل هذا القرار في تصميم البرامج التعليمية والتدريبية، وفي تشخيص الصعوبات التي يواجهها الفرد أثناء التدريب أو التعليم.

(٢) القرارات المتعلقة بالموازنة بين أسلوبي نشاط للفرد:

يُتّخذ هذا النوع من القرارات أيضا لكل فرد على حدة. ويفيد في التوجيه التعليمي، حيث تستخدم درجات الفرد في اختبارات القدرات المرتبطة ببرامج تعليمية مختلفة في توجيهه إلى البرنامج المناسب لقدراته.

(٣) القرارت المتعلقة بالموازنة بين الأفراد:

وهنا يهتم صانع القرار بالفروق بين درجات الأفراد أو المجموعات. فمثلاً يمكن أن يستخدم اختبار أو اختبارات في انتقاء أفضل الأفراد الذين يلتحقون بعمل معين، أو يحصلون على منحة دراسية معينة. ويمكن أيضا استخدام أحد المتغيرات التابعة في المقارنة بين المجموعات في إحدى التجارب.

(٤) استنتاج العلاقة بين أزواج المتغيرات:

أحيانا يهتم صانع القرار ببحث العلاقة بين أزواج المتغيرات، من مثل ذلك العلاقة بين القدرة على حل مشكلات تتضمن أشكالا هندسية، والقدرة على تكوين المفاهيم. أي أن نظرية إمكانية التعميم تراعي اختلاف طرق استخدام درجات أداة القياس نفسها باختلاف القرار المطلوب صنعه. فأحيانا يكون الاهتمام منصبا على الكشف عن الفروق الفردية (أي الموازنة بين أداء الأفراد)، وأحياناً أخرى ينصب الاهتمام على استخدام الدرجة الملاحظة لتقدير Estimate الدرجة الشاملة للفرد (أي تفسير الدرجات تفسيرا مرجعي المحك Criterion -Referenced)، أو ربما تستخدم معادلة الانحدار في تقدير الدرجة الشاملة بمعلومية الدرجة الملاحظة.

والتمييز بين دراسات إمكانية التعميم ودراسات القرارت لا يعني أن كلاً منهما منفصل عن الآخر. فالحقيقة أن التصميم التجريبي الأمثل الذي تُبنى عليه دراسة إمكانية التعميم يعتمد على القرار الذي يهم الباحث صنعه. فإذا كان الباحث مهتما بالنوع الأول من القرارت مثلا، أي القرارت المطلقة الخاصة بفرد معين دون اعتبار لمعيار الجماعة المرجعية التي ينتمي إليها الفرد، يمكنه استخدام تصميمات تجريبية بسيطة تمثل فيها معاينات مشتركة Joint Sampling من الشروط المختلفة الأبعاد أو الأوجه .Facets.

وهذا التمييز يدل فقط على أن بعض الدراسات تُجرى أثناء عملية تصميم طريقة قياس معينة، وبالتالي يمكن الاستفادة من هذه الطريقة في دراسات تالية.

فمثلا إذا أجرى باحث تحليلا عامليًا لمفردات استبانة بغرض اتخاذ قرار خاص بتجميع المفردات في أبعاد معينة على أساس تشبعاتها بهذه الأبعاد، فيمكننا اعتبار أنه يقوم بدراسة استطلاعية تشبه دراسة إمكانية التعميم G - Study.

وإذا أراد باحث آخر التوصل إلى أفضل تصميم لبطارية اختبارات تستخدم في أغراض التنبؤ، فإنه ربما يبدأ بدراسة استطلاعية يجمع فيها بيانات باستخدام بطارية مبدئية، ثم يقوم ببناء بطارية جديدة اعتماداً على معاملات الارتباط بين درجات كل اختبارين من الاختبارات التي تشتمل عليها البطارية المبدئية، وبين درجات كل اختبار ودرجات المحك، وكذلك معاملات ثبات درجات كل منها. ويمكن عندئذ استخدام هذه البطارية الجديدة في جمع بيانات خاصة بدراسات القرارات D - Studies. ومن هذا يتضح أن دراسات إمكانية التعميم (G) تُعَد جزءا من بناء أدوات القياس. ولذلك يجب أن تجرى هذه الدراسات قبل عملية جمع البيانات الخاصة بدراسات القرارت (D).

والتمييز بين دراسات إمكانية التعميم ودراسات القرارات، له أهمية خاصة فى البحوث متعددة الأوجه أو الأبعاد Multifacet Investigations. فتعدد الأوجه Facets يؤدى إلى تنوع التصميمات التجريبية التى يمكن استخدامها فى هذه البحوث. ولذلك يجب أن تكون التصميمات التجريبية المستخدمة فى دراسات إمكانية التعميم مختلفة عن التصميمات المستخدمة فى دراسات القرارات. فالباحث الذى يجرى دراسة إمكانية تعميم (G) باستخدام تصميم تجريبى معين، ويرى أنه من الممكن استخدام التصميم نفسه فى دراسة القرارات (D)، يجب أن يدرك أن باحثا آخر ربما يستخدم تصميما مختلفا للحصول على بيانات مماثلة لبياناته.

لذلك يوصى كرونباك Cronbach بأن يصاغ تقرير دراسة إمكانية التعميم بصورة تُمكِّن الباحثين المختلفين من تصميم دراسات القرارات التى يهتمون بها، وتقدير خطأ التعميم الخاص بالتصميم المستخدم.

تحديد النطاقات الشاملة فى دراسات إمكانية التعميم

Universes of generalization

إن قياس السلوك الذى نحصل عليه يكون بمثابة عينة من النطاق الشامل للقياسات الممكنة. واهتمامنا بالدرجات التى نحصل عليها من عينات من القياسات يرتبط بمدى تمثيلها للنطاق الشامل للقياسات. فلو أن الباحث استطاع قياس سمة معينة لدى الفرد بكافة الطرق الممكنة، لكانت الدرجة التى يحصل عليها الفرد فى هذه الحالة هى متوسط جميع هذه القياسات. ويطلق على هذه الدرجة فى النظرية الكلاسيكية للاختبارات «الدرجة الحقيقية True Score»، ولكن يُفَضِّل كرونباك Cronbach أن يطلق عليها «الدرجة الشاملة Universe Score» فى إطار نظرية إمكانية التعميم، لأن هذا المفهوم الجديد يتضمن عملية استدلال على النطاق الشامل للقياسات أو الملاحظات من عينة القياسات أو الملاحظات التى نجريها. كما يؤكد أن إمكانية التعميم يمكن أن تنسحب على أكثر من نطاق شامل Universe. إذ يمكن اعتبار أن الفرد عضو فى مجتمعات مختلفة Population، مثل مجتمع الأفراد الذين يقيمون فى مدينة معينة، أو مجتمع العمال، أو مجتمع الأفراد الذين يزيد دخلهم الشهرى عن مبلغ معين، أو أعضاء حزب سياسى معين ... الخ.

وكذلك يمكن اعتبار «ملاحظة» معينة أو «قياس» معين بمثابة عينة من النطاق الشامل للملاحظات أو القياسات. ولذلك إذا قلنا أن «تقدير الدرجة الشاملة هو ٧٥»، لا يكون لهذا القول معنى إلا إذا حددنا النطاق الشامل لهذه الدرجة

ويتم وصف الملاحظات أو القياسات عن طريق «الشروط أو الظروف Conditions» التى أجريت فيها، ونقصد بذلك اليوم الذى جُمعت فيه الملاحظات أو الذى أُجريت فيه القياسات، وطبيعة الموقف الذى أُجريت فيه، والشخص القائم بعملية الملاحظة أو الفرد المختبَر، وكذلك المظاهر الأخرى للعمليات التى يؤديها الفرد فى الموقف الاختبارى. وقد سبق أن ذكرنا أن المصطلح العام الذى يشير إلى شروط أو ظروف معينة يعرف باسم «البُعد أو الوجه Facet». لذلك فإن «الملاحظات» أو «القياسات» يمكن أن تُصنّف بالنسبة إلى أبعاد أو أوجه مختلفة. فهناك بُعد أو وجه يُمثل المثيرات التى تعرض على الفرد، وبُعد يمثل الوقت الذى جمعت فيه الملاحظات أو أجريت فيه القياسات، وبُعد يمثل طبيعة الموقف الاختبارى، وهكذا.

وكل هذه الأبعاد أو الأوجه منفردة أو مجتمعة تحدد النطاق الشامل، فإذا طلبنا من طفل فى يوم ما أن يرسم شجرة، فإن هذا الرسم ينتمى إلى نطاق شامل من الأشجار التى يرسمها فى مختلف الأيام، وإلى نطاق شامل من الأشجار التى يرسمها غيره من الأطفال، وإلى نطاق شامل من الأنماط المختلفة من رسومات الأطفال فى ذلك اليوم. فالاستفسار عن النطاق الشامل المطلوب إذن مرتبط بالاستفسار من الباحث عن هدفه من تفسير القياس.

ويرى كرونباك أن «النطاق الشامل للملاحظات أو القياسات» يمكن أن يتميز ببُعد أو وجه Facet واحد، أو وجهين أو أكثر، فهو يفترض أن الربط أو الجمع بين أحد شروط الوجه الأول وشرط معين من شروط كل وجه آخر، هو الذى يحدد إمكانية تعريف «ملاحظة معينة»، أو «قياس معين». ففى المثال السابق ربما يكون هناك وجهان هما: أنماط الرسم، والأطفال الذين يُطلب منهم هذا الرسم. والربط بين أى من هذه الأنماط وبين الطفل الذى يُطلب منه رسم هذا النمط هو الذى يُحدِّد تعريف «الملاحظة» المطلوبة. وإضافة وجه ثالث، وليكن الوقت الذى يُطلب فيه الرسم من الطفل، يجعل «الملاحظة» تتحدد عن طريق الربط بين النمط والطفل والموقف.

والحقيقة أن النطاق الشامل الذى تُعمم عليه «ملاحظة» معينة أو «قياس» معين يعتمد على الاعتبارات العملية أو النظرية التى تهم صانع القرار. فمثلا يختلف تقدير أحد المشرفين على العمل لأحد العمال من موقف إلى آخر. فالحالة المزاجية للمشرف وخبرته السابقة عن العامل يكون لهما تأثير على تقديره الحالى. والباحث المهتم بدراسة كفاءة العمال، يود بالطبع تعميم هذا التقدير على مجموعة التقديرات التى يمكن الحصول عليها من مواقف مختلفة، فهو يستطيع التعميم على مدة ولتكن شهرا إذا كان

التقدير عبارة عن كفاءة العامل فى نهاية العام، فأى وقت خلال هذا الشهر ربما يكون مناسبا لتقدير كفاءة هذا العامل.

أما إذا كانت البيانات المستمدة من التقدير سوف تستخدم فى إجراء دراسة للتغيرات التى تطرأ على اتجاهات المشرف نحو العمال أثناء تفاعله معهم من أسبوع إلى آخر، فإن الباحث يستطيع التعميم على يوم واحد فقط.

وإذا استُخدم التقدير كمحك يوازن فى ضوئه صدق اختبار يقيس الاستعدادات، فإن الباحث يحتاج إلى التعميم على كل من المشرفين والمواقف.

ويصبح النطاق الشامل للمشرفين غير مطلوب إذا اقتصر الهدف على دراسة التفاعل المباشر بين العامل والمشرف أثناء أدائه العمل. لذلك فإن الباحث الذى يقوم ببناء أداة قياس، أو الذى يجرى دراسة إمكانية تعميم G - Study، يجب أن يأخذ فى اعتباره بعض «الأبعاد أو الأوجه Facets»، وما يندرج تحت كل منهما من «شروط Conditions». وتسمى «الملاحظات Observations» التى نحصل عليها من مختلف تجمعات الشروط Conditions الممثلة فى دراسة إمكانية التعميم «النطاق الشامل للملاحظات المقبولة The Universe of Admissible Observations»، كما أن لكل بُعد أو وجه Facet «نطاق شامل من الشروط المقبولة The Universe of Admissible Conditions». فعند صنع القرارات استنادا إلى أسلوب قياس معين ينصب الاهتمام على تعميم هذه القرارات على «نطاق شامل من الشروط» التى نتوصل منها إلى المعلومات نفسها، ويُسمَّى هذا النطاق «النطاق الشامل للتعميم Universe of Generalization». ويمكن أن تُفيد دراسة إمكانية التعميم G - Study صانعى القرارت الفائدة المرجوة إذا كان النطاق الشامل الذى يستخدم فى دراسة إمكانية التعميم مطابقا للنطاق الشامل للشروط المقبولة أو على الأقل يكون جزءا منه.

فالنطاق الشامل للتعميم يختلف باختلاف صانع القرار. فكلما اهتم صانع القرار فى دراسة إمكانية التعميم بتحديد النطاق الشامل للملاحظات المقبولة تحديدا أكثر اتساعا بحيث يتضمن الأنشطة المحتملة للتعميم - أفادته الدراسة وأفادت غيره من المهتمين بصنع القرارت. ويتحدد النطاق الشامل للتعميم بناء على الهدف من القرار المطلوب اتخاذه.

فعند بناء أداة قياس يجب أن يجرى الباحث دراسة إمكانية تعميم G - Study تتضمن الأبعاد أو الأوجه المختلفة التى يمكن أن تفيد فى تعميم نتائج استخدام هذه الأداة. وبذلك يستطيع غيره من الباحثين تفسير هذه النتائج فى ضوء الأبعاد أو الأوجه التى يهتمون بها.



ويقترح كرونباك Cronbach ثلاث طرق مختلفة لمعالجة «بُعد أو وجه Facet» معين في دراسة إمكانية التعميم هي:

(١) تمثيل «البعد أو الوجه» عن طريق معاينة Sampling «شرطين Conditions» أو أكثر لهذا البُعد او الوجه.

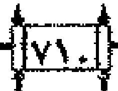
(٢) استخدام «شرط واحد ثابت Condition» للبُعد أو الوجه في جميع «الملاحظات» التي تهتم الدراسة بجمعها.

(٣) تغيير «شروط البُعد أو الوجه» دون استخدام أساليب الضبط التجريبي. ويمكن توضيح هذه الطرق بالرجوع إلى المثال السابق الخاص بدراسة تقديرات المشرفين على العمال. فربما يعتبر الباحث مواقف العمل والمشرفين عليه وجهين من النوع المذكور في الطريقة الأولى التي اقترحها كرونباك. ولذلك ربما يطلب الباحث من مشرف أو أكثر أن يقوم بتقدير كفاءة كل عامل في موقفين أو أكثر من مواقف العمل. أو ربما يعتبر أن استمارة التقدير ثابتة، أي يعتبرها «شرطا» من النوع المذكورفي الطريقة الثانية، ثم يقوم بجمع البيانات الخاصة بدراسة إمكانية التعميم باستخدام هذه الاستمارة فقط، وهذا يُمكّنه من معرفة ما سوف يحدث في حالة تغيير عبارات أو محتوى أو أشكال استمارات التقدير. وبالطبع لا يستطيع التعميم على جميع محتويات أو أشكال استمارات التقدير، كما لا يستطيع ضبط جميع الأوجه Facets الممكنة في إحدى دراسات إمكانية التعميم. ففي المثال السابق يصعب عند دراسة تقديرات المشرفين على العمال، التحكم في المدة التي يشرف فيها مشرف معين على أحد العمال. ويمكن في هذه الحالة معالجة هذه الأوجه باستخدام الطريقة الثالثة.

فصانعو القرارت يهتمون عادة بتعميم النتائج على أوجه Facets يصعب التحكم فيها، وتسهم هذه الأوجه بلا شك في أخطاء التباين عند تحليل نتائج دراسة إمكانية التعميم. ولذلك اقترح كرونباك بعض التصميمات التجريبية التي يمكن أن تستخدم في معالجة مصادر التباين التي يصعب التحكم فيها.

تحليل مكونات التباين في دراسات إمكانية التعميم:

تهدف دراسة إمكانية التعميم إلى جمع بيانات تفيد في تقدير مكونات التباين في الدرجات الملاحظة التي نحصل عليها باستخدام أسلوب معين من أساليب القياس. وقد اقترح كرونباك عددا من التصميمات التجريبية التي يمكن أن تُستخدم في تحلبا



هذه المكونات. وتختلف هذه التصميمات باختلاف عدد الأوجه والشروط التى يفترضها الباحث في دراسة إمكانية التعميم.

فإذا كان النطاق الشامل Universe يتكون من بُعد أو وجه Facet واحد، تستخدم تصميمات تجريبية أحادية الوجه Single - Facet Designs، أما إذا اشتمل على أكثر من وجه واحد، فإنه يمكن استخدام تصميمات متعددة الأوجه Multiple Facets Designs، وتقدير مكونات التباين في هذه التصميمات التجريبية يعتمد على الأساليب الإحصائية، وتختلف هذه الأساليب في درجة تعقدها وفقًا لنوع التصميم التجريبى المستخدم. وفيما يلى بعض أنواع هذه التصميمات:

(١) التصميمات أحادية الوجه Single - Facet Designs:

يشتمل «النطاق الشامل للملاحظات» فى هذا النوع من التصميمات على وجه واحد. وسوف نرمز للأفراد بالرمز (ف)، ولشروط الوجه الواحد (مفردات اختبار أو مقياس مثلا) بالرمز (ش)، ولتقاطع الأفراد مع الشروط بالعلامة (×)، ولتشابك أو تداخل الأفراد مع الشروط بالعلامة (:). وسوف نوضح بإيجاز بعض أنواع هذه التصميمات:

(أ) التصميم (ف × ش): يُعد أكثر التصميمات أحادية الوجه استخداما، وفيه يُطبق على كل فرد من الأفراد الذين عددهم (نف) مفردات عددها (نش). وهنا يكون الأفراد هم موضع القياس، حيث لا تُعدهم مصدراً من مصادر الخطأ، وبالتالى فلا يمثلون «وجه Facet»، وإمكانية التعميم تكون على جميع الصيغ الممكنة للاختبار حيث يشتمل كل منها على عينة عشوائية من مفردات مسحوبة من نطاقها الشامل.

(ب) التصميم (ش: ف): يُعد من التصميمات أحادية الوجه التى تتشابك فيه أو تتداخل الشروط مع الأفراد Nested Design. وهنا يطبق على كل فرد مجموعة مختلفة من المفردات.

(٢) التصميمات متعددة الأوجه Multiple Facets Designs:

يشتمل «النطاق الشامل للملاحظات» فى هذا النوع من التصميمات على أكثر وجه Facet واحد. وتتنوع هذه التصميمات تنوعا كبيرا، ولكنها تعتمد على المفاهيم والأسس التى تعتمد عليها التصميمات أحادية الوجه. غير أنها تتطلب تصورا فكريا أكثر عمقا، كما تتطلب عمليات جبرية ورمزية أكثر تعقيدا عند تقدير مكونات التباين فى دراسات إمكانية التعميم.



وفيما يلى بعض أنواع هذه التصميمات:

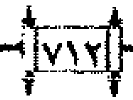
(أ) التصميم (ف × ش × ج): ويُعد من التصميمات متعددة الأوجه الشائع الاستخدام، حيث (ج) ترمز إلى المواقف، أو الاختبارات الفرعية، أو إلى القائمين بعملية التقدير. ويمكن باستخدام هذا التصميم تقدير جميع مكونات التباين ولكن كل منها مستقل عن الآخر لأن التصميم متقاطع تقاطعًا كاملا Completely Crossed، وذلك لأنه يجب أن تطبق مجموعة المفردات على جميع الأفراد فى المواقف المختلفة. لذلك يستخدم هذا التصميم فى تقدير استقرار درجات الاختبارات وتكافؤها، وذلك بتحديد النطاق الشامل لإمكانية التعميم تحديدا مناسبا. كما يمكن استخدامه فى إيجاد ثبات تقديرات المُحكمين إذا كانت (ش) ترمز إلى المحكمين.

(ب) التصميم (ف×(ش:ج): ويُعد من التصميمات المتشابكة أو المتداخلة Nested Design، حيث تتشابك أو تتداخل فيه بعض مكونات التباين أو التأثيرات Effects.

فمثلا يمكن أن تتشابك أو تتداخل مفردات الاختبار (ش) فى المواقف (ج)، أى تطبق على جميع الأفراد مجموعة المفردات نفسها فى أحد المواقف، ويطبق عليهم مجموعة أخرى من المفردات فى موقف آخر، وهكذا ويُستخدم هذا التصميم فى تقدير تكافؤ درجات الاختبارات فى إطار النظرية الكلاسيكية للاختبارات. وإذا كانت (ج) ترمز إلى أحد الاختبارات الفرعية التى تشتمل عليها بطارية من الاختبارات، فإنه يمكن استخدام هذا التصميم أيضا فى دراسة إمكانية التعميم من درجات اختبارات البطارية.

وقد تناول المؤلف هذه التصميمات المتعددة بمزيد من المناقشة فى ضوء النماذج الإحصائية فى كتابه «تطورات معاصرة فى القياس النفسى والتربوى».

ومن الجدير بالذكر أن كريك، وبرينان (Crick & Brennan, 1982) أعدا برنامجا للحاسوب أطلقا عليه GENOVA يمكن باستخدامه إجراء جميع التحليلات الإحصائية التى تتطلبها نظرية إمكانية التعميم عندما تكون التصميمات من النوع الكامل المتوازن Completely Balanced Designs. ويجرى البرنامج تحليل دراسات إمكانية التعميم G - Studies Analysis بعدد من الأوجه لا يزيد عن ستة، ونحصل باستخدامه على متوسطات مستويات الأوجه، ومتوسطات توفيقات مختلفة من مستويات الأوجه، ومجموع المربعات، ومتوسطات مجموع المربعات كما فى أساليب تحليل التباين، وقيم تقديرية لمكونات التباين، والأخطاء المعيارية لهذه القيم، ومصفوفة التباين والتغاير. كما يمكن الحصول على قيمة تقديرية لمكونات التباين، والأخطاء المعيارية



لهذه القيم في دراسة اتخاذ القرارات D - Study Anlysis استناداً إلى مكونات التباين التي يتم تقديرها في دراسة إمكانية التعميم. ويتميز هذا البرنامج بالسرعة الفائقة في تجهيز البيانات ومعالجتها حيث يستطيع معالجة ٣٥٠٠ ملاحظة في الثانية.

العلاقة بين نظرية إمكانية التعميم ونظرية الصدق:

تناولت نظرية إمكانية التعميم العلاقة بينها وبين نظرية الصدق Validity Theory بإيجاز شديد، كما ألقى كل من كامب Kamp ، وماكدونالد McDonald ضوءا خافتا على هذه العلاقة. غير أن أولى المحاولات المنظمة للتكامل بين النظريتين هي دراسة كين (Kane, 1980) ، بيد أن هذه الدراسة تحتاج إلى كثير من التطوير النظرى. فالتحقق من صدق إجراءات قياس معين يتطلب تحديد ووصف هذه الإجراءات من حيث مجموعة المثيرات التي يستجيب لها الفرد، والتعليمات التي تُعطى له، والقواعد التي يستند إليها في تقييم استجاباته. فمجموعة الشروط التي نحصل عليها من دراسة الصدق تعد واحدة من المجموعات المتعددة لهذه الشروط. لذلك ينبغي تحديد «النطاق الشامل للإجراءات المقبولة» التي تنتمي إليها هذه المجموعة من الشروط، والإفادة من دراسات إمكانية التعميم G - Studies في التحقق من صدق التكوين الفرضي للاختبارات. فالتكوينات الفرضية التي تفسر تباين الأداء في الاختبارات يمكن دراستها عن طريق تحليل مكونات التباين وفقا لأحد التصميمات المناسبة.

ويرى شيفلسون، ووب (Shavelson & Webb, 1981) أن تطوير هذا المجال سوف يكون بداية مرحلة جديدة في التنظير السيكومترى في العقود القادمة.

(ثالثا) إسهام علم النفس المعرفي في تطوير أساليب القياس والتقويم:

حدثت تطورات كبيرة في علم النفس المعرفي Cognitive Psychology الذي يعد من العلوم المتداخلة، حيث يشارك علماء النفس، وعلماء الحاسوب، والفلاسفة، وعلماء اللغة، وعلماء الأجناس البشرية، وعلماء التربية في دراسة وتحليل العمليات المعرفية الإنسانية. فمحاولة تدريس مجال تعليمي معين تستند عادة إلى نظرية أو افتراضات معينة تتعلق بالأداء الذي يتطلب عمليات معرفية عليا في هذا المجال، وكذلك إلى نظرية صريحة أو ضمنية تتعلق بعملية التعلم. فعلم النفس المعرفي المعاصر يهتم بالبحث التقييمي للأداء والتعلم، وأصبحت الإسهامات النظرية التي يقدمها تفوق إسهامات ما قدمته نظريات التعلم في العقود السابقة، كما أصبحت هذه الإسهامات أكثر ارتباطا بدراسات التعلم في الفصل المدرسي. فنظريات علم النفس

المعرفي تعد محـاولات لتفسير عمليات الفكر ، مما يتطلب نوعا من التمثيل Representation يختلف عن التمثيل الرياضي الذي يناسب العمليات الاختبارية.

ولتوضيح ذلك نفترض أنه طُلب من فرد محاولة حل بعض المشكلات اللفظية والرياضية، فعدد المشكلات التى قام بحلها يمكن التنبؤ به إذا علمنا درجات قدرته اللفظية والمكانية والعددية. غير أننا لا نستطيع باستخدام معادلة التنبؤ معرفة أي المشكلات استطاع أن يحلها فرد معين، نظرا لأن عمليات حل المشكلة هي ما ينبغى أن تتناوله النظرية المعرفية. فمن الممكن أن يجيب مجموعة من الأفراد إجابات واحدة بعمليات استدلال مختلفة .

وفحص هذه العمليات يزودنا بمعلومات عن الفرد أكثر مما نحصل عليه من مجرد تقدير درجة كلية لإجاباته. فعندما تُقدَّم مشكلة معينة فإنه يجرى وصفها داخليا فى ضوء المعلومات المختزنة بالفعل فى الذاكرة. وتشتمل الذاكرة على نوعين منفصلين من المعرفة: أحدهما معرفة تقريرية للحقائق، وثانيهما معرفة إجرائية عما يمكن فعله. وهذه المعرفة الإجرائية Procedural Knowlege تعد بمثابة «برامج» تُخبر العقل بوظائف تجهيز المعلومات المناسبة التى يمكن تنفيذها فى مواقف مختلفة. ويمكن توضيح ذلك بعملية فهم المقروء، فكثير من الطلاب يتعلمون أن الجملة أو الجملتين الافتتاحيتين فى مقال معين تحددان موضوع المقال بعامة، لذلك فإنهم يستخدمون المعلومات التى تتضمنها هاتان الجملتان لتنشيط مخطط يوجه فهمهم لبقية المقال.

وهذا يؤكد أهمية العمليات التى يستخدمها الفرد فى مواقف حل المشكلة بدلا من التنبؤ بأدائه فى هذه المواقف. وللتوصل إلى تفسير لهذه العمليات ينبغى أن يكون لدينا قدر كبير من فهم الموقف المعين، وهذا يعد أكثر صعوبة من مجرد التنبؤ بالأداء. فمثلا ينبغى معرفة سرعة التجهيز المعرفى المناسبة للموقف، والمعرفة التقريرية والإجرائية المطلوبة لحل المشكلة، وبدائل الأساليب التى يمكن باستخدامها تجهيز المعلومات المتعلقة بالمشكلة، وكيفية تحديد العمليات الأفضل لموقف معين. ومعرفة ذلك يُمكّن منظّرى علم النفس المعرفى من تناول العلاقة بين أداء الفرد وامتلاكه المعرفة المناسبة، ومدى اعتماد الأداء على وظائف معينة لتجهيز المعلومات، مثل قدرات التعميم، وتعرُّف الترميزات اللفظية، وتناول التصورات العقلية.

علاقة النظرية المعرفية بالعملية الاختبارية:

تهتم الدراسات المتعلقة بالعمليات المعرفية بوظائف تجهيز المعلومات، والمهارات العامة لحل المشكلات، أو أساليب استخدام المعرفة في مواقف محددة لحل المشكلة. غير أن دراسة التفاعل بين المستويات المختلفة للمعرفة لم يحظ إلا بقدر ضئيل من البحث مما أدى إلى عدم توظيفها في بناء الاختبارات، وذلك لأن العملية الاختبارية بطبيعتها تتناول الفرد ككل. غير أنه يمكن الإفادة من بحوث العمليات المعرفية في مجال الاختبارات، وذلك بضم المستويات الثلاثة للمعرفة التى أشرنا إليها مع الاهتمامات المتعلقة بمجال الاختبارات من حيث تطوير المقاييس المتوافرة، وبناء مقاييس جديدة، ومزيد من فهم مواقف حل المشكلة وغيرها من المواقف التى تتطلب نشاطًا معرفيا.

وقد أوضح هنت (Hunt,1985) أن كلاً من المستويات البحثية يمكن أن يسهم إسهاما معيناً فى مجال الاختبارات. فعلى مستوى تجهيز المعلومات ينبغى التميز بين تجهيز المعلومات اللغوية، وتجهيز المعلومات غير اللغوية. وتوجد تقنيات جديدة تعتمد على الحاسوب يمكن توظيفها فى بناء اختبارات تقيس وظائف تجهيز المعلومات اللغوية. فمثلا يمكن بناء مقاييس مختبرية تحدد سرعة الفرد فى تعرُّف كلمة مألوفة سواء بمفردها أو فى مضمون معين، حيث إن هذا يعد من الوظائف الأساسية فى القراءة. وتفيد هذه المقاييس فى بناء نظرية معرفية تتعلق بتفسير تباين سرعة الأفراد فى التجهيز اللغوى Language Processing. وكذلك يمكن تطوير أساليب لدراسة كيفية استجابة الأفراد لأشياء متحركة، وذلك لقياس عمليات الاستدلال البصرى المكانى، حيث إن هذا يختلف عن الاختبارات التقليدية لقياس القدرة على الإدراك المكانى، نظرا لأن الأشكال التى تشتمل عليها هذه الاختبارات تكون استاتيكية وليست ديناميكية كما فى الأساليب المقترحة. إذ يمكن بناء اختبارات تقيس قدرة الفرد على إضافة عناصر لمثير بصرى يتصوره الفرد عقليا، أو تقيس قدرته على إسقاط مسارات مثيرات متحركة للأمام بالنسبة للزمن والموقع، ويتم التحكم فى عرض هذه المثيرات وتسجيل الاستجابات باستخدام الحاسوب، وذلك لصعوبة قياس هذه القدرة باستخدام اختبارات الورقة والقلم.

كما يصعب استخدام اختبارات الورقة والقلم فى تقييم وظائف تجهيز المعلومات، ومثال ذلك قياس سرعة الفرد فى الانتقاء الإدراكى الحركى، أى يقرر مثلا أيا من المثيرين الضوئيين حدث أولاً.

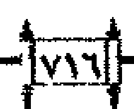
كذلك يمكن توسيع مجال الاختبارات بقياس القدرة على «توجيه الانتباه» حيث اعتبر كثير من علماء النفس هذه القدرة جزءا من الكفاءة العقلية. وأكدت البحوث المعاصرة في هذا المجال أن قدرة الفرد على توجيه انتباهه لجزء معين من مثير مركب تعد أحد المكونات الرئيسة للفروق الفردية. ويمكن قياس هذه القدرة بأن يُطلب من الفرد تحويل انتباهه من موضع معين في المجال البصرى إلى موضع آخر. كما يمكن توسيع نطاق هذه البحوث لدراسة الفروق الفردية في القدرة على تحويل الانتباه من صيغة معينة لتجهيز المعلومات إلى صيغة أخرى، ومثال ذلك التحول بسرعة من إجراء عمليات جمع إلى إجراء عمليات ضرب، وهكذا.

وبالطبع توجد أنواع أخرى متعددة من وظائف تجهيز المعلومات يمكن قياسها، ولكنها تتطلب عمليات تحكم باستخدام الحاسوب.

تصميم الاختبارات في ضوء النظرية المعرفية:

يرى كثير من علماء النفس المعرفى وخبراء القياس أن عمليات القياس وتصميم الاختبارات وعلاقتها بالعملية التربوية سوف تتخذ في القرن الحادى والعشرين توجُّهاً جديدا يختلف اختلافا جوهريا عما هو سائد في وقتنا الحاضر (Bejar et al., 1993, Hunt, 1985; Ward, 1985) . فالتطورات التى حدثت في علم النفس المعرفى، وعلم القياس النفسى والتربوى، وتقنيات المعلومات والاتصالات، وما تفرضه طبيعة هذا القرن من تحديات، تتطلب إعادة النظر في ممارسات التعليم، ونظم وأساليب القياس والتقويم لمجتمع المتعلمين.

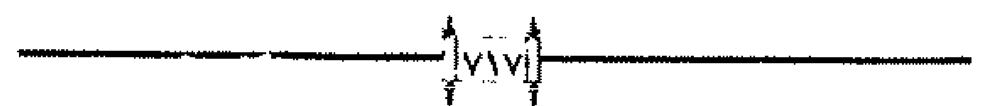
وتحقيق ذلك يتطلب توافر معلومات تختلف عما تزودنا به أساليب القياس الكلاسيكى التى تقتصر على تقديم مؤشرات عامة تتعلق بمجموعات من الأفراد بهدف المقارنة بينهم، أو التنبؤ بأدائهم الأكاديمى. فعملية التقويم سوف تصبح جزءا لا يتجزأ من عملية التعليم، أى أنهما سيصبحان عملية واحدة متكاملة. فالاختبارات الحالية التى تستخدم فقط في أغراض انتقاء الطلاب للالتحاق بمراحل التعليم المتتالية، ولا تُقدم معلومات يمكن الاستفادة منها في التقدم التربوى سوف تكون قليلة الجدوى، وذلك لأننا لا نستطيع باستخدامها قياس نمو وتطور الكفايات الإنسانية. وهذا يتطلب التحقق من الفروق الجوهرية بين الأداء المتميز والأداء الضعيف عن طريق تقييم بنية المعرفة والعمليات المعرفية التى تكشف عن درجات كفايات الفرد في مجال دراسى معين.



فمعظم تقنيات القياس في وقتنا الحاضر صُممت لكي تُستخدم بعد بناء مفردات الاختبارات. فقد لاحظنا في الأبواب السابقة أن تحليل مفردات الاختبارات، والتحقق من صدقها وثبات درجاتها ، وإيجاد معاييرها، وتحليل أبعاد الاختبار ومكوناته العاملية يتم بعد بناء المفردات. بينما تشير التوجهات المعاصرة إلى أنه ينبغي الاهتمام بإسناد تصميم المفردات إلى نظريات معرفية تتعلق بخصائص الكفاءة المرجوة، والبُنى والعمليات التي تكتسب وتنمو لدى الفرد المبتدئ حتى يصبح لديه كفاءة عالية نتيجة تعلمه معرفة واستراتيجيات متقدمة. وبذلك يمكن باستخدام هذه الاختبارات تحديد إنجازات الفرد في مختلف مستويات الاكتساب. وهنا لا يكون التركيز منصبا على المحتوى وإنما على البنى والعمليات المعرفية التي تيسر اكتساب ونمو الكفاية المرجوة (Messick, 1984) . فالنظريات السيكومترية المتعارف عليها ركزت على قياس عمليات عامة، فالقدرات العقلية تقاس في عموميتها، مثل القدرة اللفظية أو العددية أو المكانية، وكذلك الاستعدادات الخاصة بأنواعها المختلفة التي يتبين ارتباطها بالنجاح الأكاديمي. كما أن نظريات التعلم المتعارف عليها هدفت لتقديم أدلة تتعلق بالعمليات والشروط العامة للتعلم، والقوانين التي تفسر أنماط التعلم المختلفة، مثل أنواع الاشتراط، وطبيعة التعزيز، والتغذية الراجعة نتيجة التعلم، والممارسة الموزعة والمركزة.

وعلى الرغم من أن هذه الأدلة قد أسهمت في تفسير مبادئ عامة مهمة في التعلم الإنساني، إلا أن جليزر (Glaser, 1984) أوضح أن الاهتمام أصبح موجها نحو دراسة أنماط معينة من المعرفة والمهارات التي يكتسبها الأفراد ويواجهونها في سياق حياتهم. وهذا التوجه الجديد يؤكد التعلم في نطاقات معرفة ثرية تتعلق بالمجالات الدراسية والأنظمة الأكاديمية مما يمكننا من تحديد جوانب القوة والضعف في أداء المتعلمين لمهام أكاديمية. فبدلا من محاولة تحديد جوانب ضعف عامة لديهم يتم توجيه الاهتمام إلى معاونة كل منهم في التعرف بدقة على ما لديه من معرفة جزئية أو غير مكتملة لكي تحاول عملية التعليم التركيز عليها مباشرة.

وأشار جليزر إلى بعض التطورات التي حدثت في هذا الاتجاه، من بينها تحليل الأداء استنادا إلى أساليب معالجة المعلومات للإسهام في تنمية كفاءة المتعلمين في المهام الأكاديمية، وكذلك الفهم المستبصر لطبيعة الأداء الكفء نتيجة دراسة خصائص الفرد الكفء والفرد المبتدئ في النطاقات المختلفة للسلوك الإنساني. ومثال ذلك اكتساب الأطفال الحقائق الحسابية، والمعرفة والأساليب التي يتبعونها في حل المسائل الكلامية، والمبادئ التي ينطوي عليها مفهوم العلامات العشرية، وغير ذلك. وهذه



الدراسات المتعلقة بتشخيص وتصنيف أنماط الاخطاء في أداء الأطفال في الحساب تعد أساسا للتشخيص الدقيق لفهم الطفل المبتدئ في تعلم الحساب وكذلك سوء فهمه. وأوضح ريزنك (Resnick, 1984) أن هذا ييسر تحديد المعرفة والإجراءات غير المكتملة التي تسهم في ضعف أداء الأطفال في الحساب بحيث يمكن علاج هذا الضعف أثناء عملية التعليم.

فالأطفال الذين يخطئون في إجراء العمليات الحسابية يفترض أنهم لم يتبعوا إجراءات معينة بدقة وكفاءة. لذلك فإن علم النفس المعرفي يمكن أن يزيد من فاعلية عملية التعليم ليس فقط ببحث إجابات الطفل عن مجموعة من المشكلات، وإنما أيضا ببحث العمليات التي توصل عن طريقها إلى الإجابة بحيث يمكنه تعلم العمليات الصحيحة. ونظام تقدير الدرجات الذي يحدد الأخطاء المتظمة له انعكاساته في العملية الاختبارية، وذلك لأن تقويم الأطفال لا يتم على أساس عدد الأخطاء في إجابتهم في الاختبارات، وإنما على أساس سوء فهمهم أو استخدامهم قواعد غير مكتملة مما يؤثر في أدائهم. فتشخيص الأداء بهذا الشكل يربط بين عملية التقويم وعملية التعليم.

كذلك صمم بعض الباحثين نماذج رياضية تتعلق بكيفية تعلم الأطفال مهارات القراءة والكتابة استنادا إلى مقارنات بين الأساليب والإجراءات التي يتبعها الأطفال المبتدئين، والأطفال الأكثر تقدما. وأوضحت الدراسات التي تناولت الأداء الكفء أن قدرة الإنسان على أداء كثير من المهام التي تتطلب الانتباه تكون محدوده. فإذا كان التجهيز الآني Simultaneous Processing لهذه المهام التي يتضمنها نشاط معقد يتطلب انتباهًا يقظًا، فإنه تكون هناك صعوبات نظرا لأن الانتباه ينبغي أن يتحول من مهمة إلى أخرى. ولكن إذا أصبح أداء بعض هذه المهام آليا من خلال الممارسة ولا يتطلب عندئذ انتباهًا يقظًا، فإن الفرد يمكن أن يكرس جهده لأداء المهام الأعلى (Glaser, 1984; Greeno, 1984; Kintsch, 1985) ويتضح هذا التناسق بين مكونات المهام عند دراسة مهارات القراءة، وبخاصة العلاقات بين مهارات القراءة المتعلقة بالكلمات، وكذلك عمليات فهم المقروء. فانتباه القارئ يتحول بين مهارات فك الترميز المتعلق بتعرف الكلمات،ومهارات فهم المقروء التي تعمل على تكامل الأفكار التي يتضمنها موضوع القراءة في الذاكرة. ويمكن ملاحظة التحول في الانتباه إذا كان القارئ مبتدئا حيث يكون تركيزه متبادلا بين معرفة نطق الكلمة وبين معنى الكلمة في سياق المقروء والانتباه لكل من هذه العمليات يستغرق وقتا من القارئ، وبطء تفاعلها مع

العمليات الأخرى يمكن أن يؤدى إلى عدم الكفاءة فى القراءة. لذلك فإن الأداء المنخفض فى القراءة يعكس فى كثير من الأحيان تداخل تأثير بطء وعدم فاعلية فك الترميز فى تنفيذ فهم مهام من مستويات أعلى. وهذا التداخل له تضميناته فى تقييم التعلم.

فبعض العمليات تتطلب وصولها إلى مستوى معين من الكفاءة لدى الفرد بحيث يمكن تنفيذ عمليات أخرى معا وبطريقة متسقة. ولكى يكون التعلم فاعلا ينبغى تقييم مستوى فاعلية المهارات الأساسية المطلوبة لتقليل تداخل هذه التأثيرات مع العمليات التى من مستوى أعلى، بحيث تيسر كل عملية العملية الأخرى. وهذا يتطلب تصميم أساليب تشخيص الكفاية فى المهارات الأساسية بحيث يتبين نجاحها فى تركيز الانتباه للمستويات العليا. ويبذل علماء النفس المعرفى جهودا كبيرة فى محاولة فهم البنية المعرفية والقدرات لدى الأفراد الذين يتميزون بمهارة معينة، وتحليل العمليات المتضمنة فى تحويل الأفراد المبتدئين إلى متميزين تدريجيا. ففهم طبيعة التميز أو الكفاءة سوف يمكن خبراء القياس من تطوير أساليب لتقييم الأداء فى مستويات كفاءة متعددة.

وقد أوضحت نتائج الدراسات فى هذا المجال أن الأفراد المتميزين يعملون على تنمية تنظيمات للمعرفة تمكنهم من إدراك أنماط ذات معنى بسرعة فى ذاكرتهم وهذا يسمح لهم بتكوين تمثيلات Representations للمشكلات تمكنهم من التصرف الجيد المناسب. بينما يقوم الأفراد المبتدئون بتمثيل المشكلات بطريقة مختلفة اختلافا نوعيا واصطناعيا مما يجعل المواقف المشكلة أكثر صعوبة فى حلها.

فتعرُّف الأنماط Pattern Recogition، وتمثيل المشكلات Problem Representation تعد بمثابة مؤشرات للكفاءة، وينبغى تضمينها فى تقييم نمو مهارات المبتدئين ليصبحوا متميزين فى مجال معين، وكذلك ينبغى الاستناد إليها كأساس لبناء مفردات الاختبارات قبل إجراء العمليات السيكومترية، مثل تحليل المفردات أو تقدير الثبات. ولعل هذا يعد من التوجهات المهمة التى ينبغى أن تنال قدرا وافيا من البحوث المستقبلية بحيث لا يعتمد تقييم التعلم أو الأداء على مجرد درجة كلية أو تقدير عام أو مئينى، وإنما يعتمد على نظام تقدير تعليمى Instructional Scoring يوجه التعلم ويوضح متطلبات رفع مستوى الكفاءة لدى المتعلمين.

خلاصة

تناولنا في هذا الفصل بعض أهم التطورات المعاصرة في مجال القياس والتقويم التربوى والنفسى التى أصبح خبراء القياس يركزون عليها اهتمامهم ويولونها مزيدا من البحث والدراسة، وتنعقد حولها العديد من المؤتمرات المتخصصة فى كثير من الدول المتطورة. ويتعلق أحد هذه التطورات بالنظرية السيكومترية المعاصرة التى يطلق عليها نظرية الاستجابة للمفردة الاختبارية IRT وما تتضمنه من نماذج، مثل نموذج البارامتر الواحد (نموذج راش)، والنموذج ثنائى البارامتر (نموذج لورد)، والنموذج ثلاثى البارامتر (نموذج بيرنبوم)، وغيرها من النماذج التى انبثقت عنها .

وتحاول هذه النماذج التغلب على أوجه قصور النظرية الكلاسيكية للاختبارات التى سيطرت على الفكر التربوى والنفسى ما ينوف عن نصف قرن، وقد أوضحنا فى هذا الفصل بعض أهم هذه الأوجه. فهذه النماذج الجديدة تركز على المفردة الاختبارية وتأثيرها فى الاختبار ككل. وكل منها يعبر عن التوزيع الاحتمالى للاستجابات الممكنة للفرد لمفردة اختبار معينة، وتشتق خصائص الاختبار ككل من خصائص المفردات التى يشتمل عليها، على عكس النظرية الكلاسيكية التى تركز على الاختبار بعامة. لذلك تتميز هذه النماذج بميزات متعددة، فإذا تم تعيير مجموعة من المفردات (أى تقدير بارامتراتها) وفقا لنموذج راش مثلا، فإنه يمكن بسهولة تحويل درجات الاختبارات التى يشتمل كل منها على مجموعة فرعية من المجموعة الأصلية إلى ميزان مشترك Common Scale وهذا ييسر عملية تكافؤ الاختبارات Test Equating، وبناء اختبارات مفصلة أى موائمة لمستوى قدرة الفرد الذى يُختبر بها كما سنوضح فى الفصل التالى. كما أن هذه النماذج تتميز بأنها تزودنا بأساس إحصائى سيكومترى يستند إليه فى تصميم الاختبارات.

والتطور الثانى يتعلق بنظرية الثبات متعدد الأبعاد التى يطلق عليها نظرية إمكانية التعميم Generalizability Theory . واهتمت هذه النظرية أيضا بمعالجة بعض أوجه قصور النظرية الكلاسيكية للاختبارات فيما يتعلق بمفهومى الثبات والصدق، وتوسيع نطاق هذه النظرية من أجل تعميق وتفسير كل من هذين المفهومين، ومعالجتهما معالجة أكثر شمولية. فهذه النظرية المعاصرة تسمح بمعالجة مصادر متعددة للخطأ،

كما تسمح بتعريفات متعددة للدرجات الحقيقية. فالنظرية الكلاسيكية لم تميز بين مصادر الخطأ المحتملة فى هذه الدرجات الحقيقية. لذلك تهتم نظرية إمكانية التعميم بمدى تعميم عينة القياسات المستمدة من مجموعة من البيانات أو الملاحظات على المجموعة الشاملة لهذه البيانات. وتحاول دراسات إمكانية التعميم G-Studies تحديد تأثير معاينات شروط أو ظروف مختلفة على درجات الاختبارات من أوجه Facets مختلفة، وتقدير مكونات التباين التى ترجع إلى اختلاف هذه الأوجه، وكذلك تقدير تفاعل هذه المكونات باستخدام أساليب تحليل التباين Analysis of Variance وفقا لتصميمات مناسبة أشرنا إليها. وبذلك يمكن الحصول على معامل إمكانية التعميم الذى يدل على النسبة بين تباين الدرجة الشاملة إلى تباين الدرجة الملاحظة،ويُعد هذا المعامل بديلا عن معامل الثبات الذى تتناوله النظرية الكلاسيكية للاختبارات. وكذلك يمكن الإفادة من دراسات القرارات D - Studies فى جمع بيانات تفيد فى اتخاذ العديد من القرارات، والتوصل إلى تفسيرات واستنتاجات بناء على نتائج دراسات إمكانية التعميم D - Studies أوضحنا بعضا منها بأمثلة، كما أوضحنا علاقة هذه الدراسات بنظرية الصدق.

أما التطور الثالث فيتعلق بإسهامات علم النفس المعرفى فى تصميم الاختبارات بحيث يتحول الاهتمام من استخدام الاختبارات فى تقويم التحصيل العام للمتعلمين فقط إلى الإفادة من البيانات المستمدة منها فى تحسين وإثراء عملية التعليم وتقويم المعلمين والأساليب التعليمية. وقد أدى هذا التوجه إلى توظيف نتائج بحوث علم النفس المعرفى فى تصميم الاختبارات من منظور جديد يركز على العمليات المعرفية المتضمنة فى الأداء الكفء فى مختلف المجالات الدراسية. وسوف يؤدى ذلك إلى بناء اختبارات تختلف اختلافا ملحوظا عن الاختبارات المستخدمة فى وقتنا الحاضر من حيث التصميم والشكل. فالجيل الجديد من الاختبارات سوف يركز على قياس عمليات تجهيز المعلومات ومعالجتها وما تتطلبه من انتباه يقظ، وترميز، وفك الترميز، وتعرُّف الأنماط، وتمثيلات معرفية مختلفة، وغير ذلك من العمليات التى يستخدمها المتعلم فى إجابته عن مفردات الاختبار. وسوف يساعد ذلك فى تنمية مهارات المبتدئين لكى يصبحوا مقتدرين فى مجال معين من مجالات المعرفة استنادا إلى أساليب علمية، وتقويم ذلك تقويما تشخيصيا متضمنا فى عملية التعليم ومتكاملا معها. وسوف يؤدى ذلك إلى تغيرات جوهرية فى نظريات القياس وممارسات التقويم.

الفصل الخامس عشر

بعض التوجّهات التطبيقية

* مقدمة
* تطوير أساليب قياس وتقويم التحصيل الدراسى
  (أولا) استخدام الاختبارات مرجعية المحك فى تقويم التحصيل
  (ثانيا) تقنيات تفريد قياس التحصيل باستخدام الاختبارات المفصلة
* الاستراتيجية ذات المرحلتين
* الاستراتيجية متعددة المراحل
* نموذج التفريغ الثابت
* نموذج التفريغ المتغير
* استراتيجية الاستعانة بالحاسوب
  (ثالثا) بناء نظم بنوك الأسئلة
* مراحل إنشاء نظام بنك أسئلة مقترح
* البنك الدولى للأسئلة، وتقويم التحصيل التربوى من منظور دولى
  (رابعا) التقييم التوثيقى الشامل للتحصيل والكفايات
* مكونات الصحائف الوثائقية Portfolios
* صيغ الصحائف الوثائقية
* بعض مشكلات تقييم التحصيل باستخدام الصحائف الوثائقية
* خلاصة

مقدمة

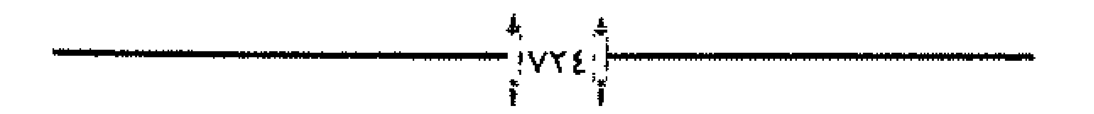
أسهمت التطورات النظرية السيكومترية التى أوضحناها بإيجاز فى الفصل السابق وغيرها من التطورات المعاصرة فى إثراء تطبيقات القياس والتقويم التربوى والنفسى وتوسيع نطاقها فى مجالات متعددة. وعلى الرغم من أن بعض هذه الإسهامات لا تزال فى مرحلة التجريب، والبعض الآخر أصبح واقعا مستخدما فى تصميم وبناء العديد من الاختبارات والمقاييس المنشورة فى وقتنا الحاضر، إلا أن هناك جهودا متزايدة من جانب خبراء القياس لتذليل العقبات التى تحول دون الإفادة المثلى من هذه الإسهامات فى مختلف الجوانب العملية والتطبيقية، وتبسيطها للممارسين والدارسين فى هذا المجال. فتصميم الاختبارات والمقاييس من منظور جديد أصبح يستند إلى التطورات التى حدثت فى علم النفس المعرفى، وما أسهمت به فى إلقاء مزيد من الضوء على العمليات المعرفية التى ستؤدى إلى توجه جديد فى تصميم عملية التعليم والأنظمة الاختبارية.

كما أن التطورات التقنية وبخاصة تقنيات الحاسوب منخفضة الكلفة سوف تلعب دورا مهما فى الصناعة المتعلقة بالاختبارات Testing Industry . وسوف يكون لذلك انعكاساته الواضحة على نظم وأساليب القياس والتقويم التربوى والنفسى فى هذا القرن.

لذلك سوف نتناول فى هذا الفصل بعض أهم التوجُّهات التطبيقية، والقضايا المتعلقة باستخدام الاختبارات والمقاييس من هذا المنظور الجديد.

تطوير أساليب قياس وتقويم التحصيل الدراسى:

ينظر الكثيرون إلى التحصيل الدراسى نظرة تقليدية محدودة، حيث يرون أن تحصيل الطلاب يتعلق ببعض الجوانب المعرفية التى يكتسبونها من تعلم المحتوى الدراسى المقرر والتى يسهل على المعلم قياس وتقويم مدى تحققها. وفى إطار هذا المفهوم الضيق أصبحت تقتصر الاختبارات التحصيلية الصفية والعامة على قياس قدرة الطلاب على استرجاع الحقائق والمعلومات المتفرقة المتعلقة بالمحتوى الدراسى باستخدام أسئلة ومفردات اختبارية بسيطة. وتبنى هذه الاختبارات فى غالبية الأحيان دون الاستناد إلى نظرية أو نموذج متطور فى القياس التربوى، وتفسر درجاتها دون



تحديد واضح للإطار المرجعي الذى يعطى معنى ودلالة لهذه الدرجات، وبذلك لا تشكّل هذه الأسئلة والمفردات فى مجملها اختبارا تحصيليا بمفهومه العلمى، نظرا لأنها لاتمثل النطاق السلوكى الشامل للمعارف والمهارات التى ينطوى عليها المحتوى الدراسى، إذ أن إجابتها لا تتطلب عادة قدرا كافيا من التفكير، أو إدراك العلاقات، أو تقييم الأدلة، أو اقتراح بدائل وتوجهات، أو اتخاذ القرارات، أو حل المشكلات التى يمكن أن يواجهها الفرد المتعلم فى تعامله مع معطيات حياته ومطالبها الملحة. فكيف والحال كذلك يمكن الإفادة من نتائج مثل هذه الاختبارات التحصيلية والاستناد إليها فى اتخاذ قرارات صائبة تتعلق بالمتعلم كإنسان مفكر ومبدع، وبالمعلم كمصمم ومهندس تعليمى، وباستراتيجيات التعليم كمنهجية وأسلوب علمى، وبالبرنامج التعليمى كمنظومة متفاعلة ومتجددة.

فالاختبارات التحصيلية بوضعها الراهن تُبنى على فلسفة تربوية تؤكد بل وتشجع إبراز الفروق الفردية، وتحث على التعليم التنافسى من أجل حصول الفرد على مركز نسبى متفوق بين أقرانه دون محاولة تحديد ما يمتلكه الفرد من مهارات وظيفية وأخلاقيات وسلوكيات إيجابية بناءة أو أخذ المصالح المشتركة للجماعة كفريق ينبغى أن يعمل متآلفا لخير المجتمع ورقيه بعين الاعتبار.

فهذه النظرة الضيقة والمحدودة لمفهوم التحصيل الدراسى والفلسفة التربوية التى تهتم بالفروق بين الأفراد والموازنة بينهم وما ترتب على ذلك من أساليب قياس وتقويم للتحصيل تركز على ما اختزنه المتعلم فى ذهنه من معلومات محددة لم تعد تناسب المتطلبات المستقبلية للتربية واحتياجاتها المتغيرة فى هذا القرن الذى يتميز بالتفجر المعرفى، والتقدم التكنولوجى، وثورة المعلومات والاتصالات. فالتربية سوف تقود التغيير وتكون فى مقدمته، وسوف تلعب دورا رئيسا فى تنمية المجتمعات المستقبلية التى نطمح إليها. كما أن القضايا والمشكلات التربوية الحيوية المتعلقة بالتحديات الثلاثة المرتبطة والمتفاعلة التى تواجهها البشرية وهى: التنمية، والحفاظ على البيئة وحمايتها، والسلام سوف تصبح المحور الرئيسى للسياسات الدولية والقومية، هذا بالإضافة إلى أن المتغيرات المتعلقة بالنمو السكانى، والتطورات العلمية والتكنولوجية، وأنماط الوظائف والأعمال المستحدثة يُتوقع أن تصبح من متغيرات التربية والتعليم مستقبلا (Klchinkoff, 1922; Naisbitt 1984; UNESCO, 1989; USIA, 1991)..

ففى إطار هذه الأهداف المستقبلية للتربية وفى ضوء هذا المنظور الرحب للتحصيل الدراسى لا تكون عملية التقويم غاية فى ذاتها لتحديد نجاح المتعلمين

وانتقالهم إلى الصفوف الأعلى أو المراحل التعليمية اللاحقة أو رسوبهم، وإنما تصبح جزءا متكاملا من عملية التعليم توجهها وتعززها وتصحح مسارها بما يحقق هذه الأهداف. وهذا يتطلب التحول من أساليب ونظم الامتحانات التقليدية السائدة والدرجات والتقديرات الاعتبارية التي تقتصر على الموازنة بين أداء المتعلم وأداء أقرانه في اختبارات تهتم بإبراز الفروق الفردية وتؤكدها إلى أساليب ونظم تؤكد الشخصية المتكاملة والمتوازنة للمتعلم وما يمتلكه من كفايات ومهارات وظيفية، والفهم العميق والمتسع لمضمون المواد الدراسية الذي يكتسبه المتعلم من خلال أساليب التعلم الذاتي وحب الاستطلاع بما يُمكنه من التعامل مع بيئته وإثرائها.

وقد توصل كروكس (Crooks, 1988) في دراسة شاملة حول أثر التقويم داخل الصف على المتعلمين إلى وظائف متعددة لعملية تقويم التحصيل الدراسي على مستوى المدرس، ومستوى المقرر، وعلى المدى الأبعد. ويمكن إيجاز بعض الوظائف فيما يلي:

على مستوى الدرس:

- تنشيط وتعزيز المعارف والمهارات اللازمة للبدء في دراسة موضوع دراسي جديد.

- تركيز الانتباه على الجوانب المهمة في موضوع الدرس.

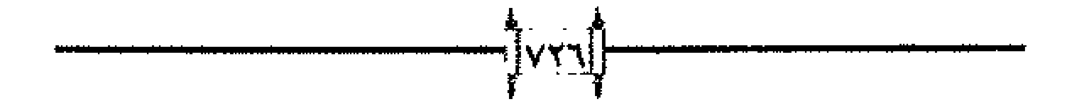
- حفز المتعلمين على استخدام استراتيجيات تعلم نشطة.

- إتاحة الفرص الكافية للمتعلمين للتدرب على المهارات تدربا كافيا يسمح بتعميق التعلم.

- تعريف المتعلمين بنتائج جهودهم وإرشادهم إلى كيفية توجيه هذه الجهود وتصحيح مسارها.

- إرشاد المتعلمين إلى كيفية متابعة تقدمهم الدراسي بأنفسهم وتنمية مهاراتهم في التقويم الذاتي بما يؤهلهم لانتقاء الأنشطة التعليمية والتعلمية التالية من أجل زيادة درجة تمكنهم واتقانهم.

- تنمية حس الإنجاز لدى المتعلمين.



على مستوى المقرر أو الموديول التعليمي:

ـ التأكد من تحقيق المتعلمين للمعارف والمهارات المطلوبة للتعلم الفاعل للمقررات الدراسية.

ـ حث المتعلمين على دراسة المقرر وتنمية إدراكهم لقدراتهم فى دراسة موضوعاته.

ـ تعريف المتعلمين بالأهداف الأساسية المتكاملة للمقرر، ومستويات الأداء المرجو تحقيقها.

ـ تشجيع المتعلمين على انتقاء وتطوير استراتيجيات وأساليب استذكار تثرى تعلمهم.

ـ تحديد إنجازات المتعلمين فى المقرر بما يسهم فى نشاطاتهم المستقبلية.

على المدى الأبعد:

ـ تعزيز قدرة المتعلمين على استبقاء المعلومات المتعلمة، وتوظيفها فى مواقف متنوعة وبطرق متعددة.

ـ تعزيز مهارات التعلم وأساليب الاستذكار النشط لدى المتعلمين.

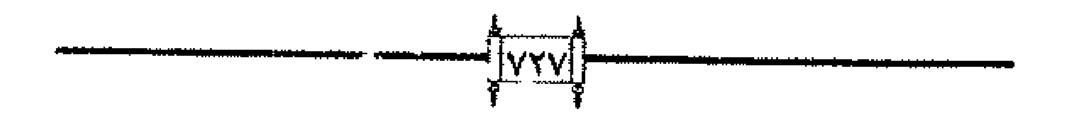
ـ الحفز المستمر لدافعية المتعلمين من أجل استمرارية التعلم.

ـ تنمية إدراك المتعلمين لجوانب التميُّز فى شخصياتهم وكفاءاتهم.

وتحقيق هذه الوظائف يتطلب ضرورة الأخذ بالمفهوم العلمى لعملية التقويم فى إطارها الشامل كما أوضحناه فى الفصل الأول باعتبارها عملية منهجية تتطلب جمع بيانات موضوعية ومعلومات صادقة باستخدام أدوات مقياس متنوعة فى إطار مجموعة من الأهداف المحددة للوصول إلى تقديرات كمية وأدلة كيفية يُستند إليها فى صنع القرارات وإصدار الأحكام المناسبة التى تتعلق بالمتعلمين وبعملية التعليم، وذلك لتحسين نوعية الأداء ورفع درجة الكفاءة بما يساعد على تحقيق هذه الأهداف.

(أولاً): استخدام الإختبارات مرجعية المحك فى تقويم التحصيل
:Criterion - Referenced Evaluation

يتضح مما سبق أن الأساليب السيكومترية الكلاسيكية فى القياس التربوى وما تستند إليه من فلسفة تربوية تحث على التنافس الفردى من أجل التفوق على الأقران



تجعل عملية تقويم التحصيل محدودة الفائدة وذات تأثير سلبى على المتعلمين والمناخ المدرسى، وبذلك لا تسهم فى تحقيق عملية التقويم لوظائفها المرجوة التى أشرنا إليها.

فالاختبارات التحصيلية التى تُبنى وفق هذا التوجه التنافسى تستهدف التمييز والتفاضل بين الطلاب من أجل وضع تقديرات لهم فى نهاية المدة الدراسية أو اتخاذ قرارات الانتقاء. لذلك يحاول الطالب جاهدا توقع الأسئلة التى سوف يشتمل عليها الاختبار، واسترجاع ما اختزنه فى ذاكرته من حقائق وتعميمات، والعمل منفردا بغية النجاح.

ويفترض فى هذا التوجه أن تعلم الطالب يعتمد اعتمادا أساسيا على قدرته والجهد الذى يبذله، وعلى العكس من ذلك يُفترض أن إخفاق الطالب يرجع إلى انخفاض قدرته أو سلبية اتجاهه نحو التعلم أو المادة الدراسية. أى أن الفروق الفردية فى إطار هذا التوجه تعد بمثابة معطيات. وبذلك يحظى الطالب الأكثر قدرة على رتبة أعلى من الطالب الأقل قدرة. غير أن هذا يترتب عليه آثار ضارة بكل منهما، فالطلاب الناجحون يجدون أن التعليم الذى يتلقونه يغفل تنمية مهاراتهم فى التفكير وحل المشكلات التى يمكن أن يواجهونها. أما بالنسبة للطلاب الراسبين فالضرر أكبر حيث إنهم يجدون أن التعليم الذى يتلقونه غير مناسب لهم وربما يحرمون من فرص متابعة دراستهم.

ومما يؤخذ على هذا النوع من الاختبارات التحصيلية التى يُطلق عليها «الاختبارات مرجعية الجماعة أو المعيار "Norm - Referenced Tests" أن التأكيد الأساسى فيها يكون على الدرجات الكلية، وبذلك يقل الاهتمام فيها على دقائق المحتوى بينما يزيد التركيز على التجريدات الأكثر عمومية. ولهذا فإن الاختبار يحدد الدرجة الكلية فى العمليات الحسابية مثلا ولكنه لا يوضح شيئا عن نوع العملية الحسابية التى يستطيع الطالب أو لا يستطيع إجراءها. فالدرجة الكلية يصعب استخدامها فى تحديد مواطن الضعف أو الأخطاء الشائعة أو فجوات التحصيل. لذلك لا نعجب إذا وجدنا أن المعلمين أو الطلاب لا يستفيدون شيئا من نتائج هذه الاختبارات التحصيلية سواء الصفية أو المقننة فى إثراء عملية التعليم والتعلم. كما أن انتقاء مفردات هذه الاختبارات يعتمد بدرجة كبيرة على الأساليب الإحصائية فى تحليل مفرداتها حيث تستبقى فقط المفردات التى تميز بدرجة أفضل بين الطلاب وتسمح بترتيبهم بالنسبة لبعضهم البعض.. لذلك فإن الاختبار فى صورته النهائية يكون مشتملا على مفردات

متوسطة الصعوبة وترتبط درجاتها ارتباطا مرتفعا بالدرجة الكلية فى الاختبار. وانتقاء المفردات التى يتكون منها الاختبار التحصيلى على هذا الأساس يقلل من صدق محتوى الاختبار، أى مدى تمثيل مفرداته لنطاق المحتوى الذى يقيسه، ويحيد بالاختبار عن الغرض منه كأداة قياس نظرا لما يعترى نطاق المحتوى من تغيرات غير معلومة أو محددة أثناء عملية تحليل المفردات. لذلك فإن إضافة أو حذف المفردات بهذا الشكل من أجل التحكم فى مستوى صعوبتها وتمييزها يؤثر تأثيرا بالغا فى صدق الاختبار. هذا بالإضافة إلى أن كثرة عدد مثل هذه المفردات التمييزية تجعل الاختبار أقرب إلى اختبارات الاستعدادات منه إلى اختبارات التحصيل. وعندئذ لا يصبح الاختبار أداة قياس للتحصيل، وإنما أداة للتمييز بين الطلاب. فقياس التحصيل ربما يؤدى إلى تمييز جيد بين الطلاب أو ربما لا يؤدى إلى ذلك (Izard, 1990). فالهدف من قياس التحصيل كما حدده جليزر (Glaser, 1963) بوضوح هو «تقديم معلومات تفصيلية محددة عما يستطيع المتعلم أداءه وما لا يستطيع». فعملية قياس تحصيل الطلاب إذن تعد عملية تحليلية تقنية، بينما عملية التمييز بينهم تعد عملية سيكومترية. فالاختبار الذى يصمم بهدف قياس التحصيل ينبغى أن يكون تركيزه الأساسى على محتوى المفردات وليس على خصائصها الإحصائية التى تناسب التوزيع الاعتدالى. فاختبار التحصيل يهدف لتقرير مستوى أداء الفرد بالنسبة لنطاق سلوكى معرف تعريفا جيدا ومحددا وليس مجرد توضيح أن الطالب قد حصل على درجة كلية أعلى أو أقل من معظم أقرانه. وتصبح الخطوة الأساسية فى بناء اختبارات التحصيل من هذا المنظور هى تحديد محتوى المفردات تحديدًا تفصيليًا واضحًا، بل وينبغى أن يكون نطاق المحتوى كاملا بقدر المستطاع.

ونعنى بذلك أنه ينبغى حصر جميع مفردات النطاق الممكنة صراحة أو ضمنيا. وهذا يتطلب تحديد مواصفات النطاق المعين بدلا من الاعتماد على جدول مواصفات الاختبار الشائع الاستخدام. فإذا ما تم تحديد نطاق المحتوى Content Domain تحديدا كاملا لا تصبح عملية تحليل المفردات غير ضرورية فحسب بل وغير مرغوبة وذلك من أجل الحفاظ على صدق معاينات هذا النطاق Content Sampling Validity، ويكون الحكم على نوعية الاختبار كمقياس للتحصيل فى ضوء نطاق المحتوى وليس لكل مفردة على حدة. غير أن التحديد الكامل لنطاق المحتوى يكون ممكنا بدرجة أكبر فى الجوانب النفسية الحركية وبعض المجالات الدراسية كالرياضيات والعلوم، أما فى المجالات الوجدانية ومجالات العلوم الإنسانية ربما يصعب ذلك، ويقتصر على تحديد الأهداف أو النواتج السلوكية المرجوة من دراسة المحتوى، ووضع مواصفات تفصيلية

لهذه الأهداف، وما يجب أن تشتمل عليه مفردات الاختبار. فعندئذ يمكن استخدام أساليب تحليل المفردات فى التعرف على المفردات الجيدة والمفردات التى تحتاج إلى مراجعة. غير أنه فى هذه الحالة لا ينبغى استبعاد المفردات التى يتبين أن معامل تمييزها منخفض أو سلبى وإنما يجرى فحصها لمعرفة جوانب القصور فيها (Berk, 1978; Popham, 1980).

ومما لا شك فيه أن التركيز على نطاق المحتوى وخصائصه وليس على الخصائص الإحصائية لمفردات الاختبار يتطلب إعادة النظر فى الممارسات السائدة فى تصميم وبناء الاختبارات التحصيلية الصفية والعامة والفلسفة التربوية التى تستند إليها هذه الممارسات. فنطاق المحتوى يصبح الإطار المرجعى أو المحك Criterion الذى تُنسب إليه درجات الاختبار وليس معيار الجماعة Norm كما هو الحال فى الاختبارات التحصيلية الكلاسيكية. وبذلك لا يقتصر تقرير نتائج الطلاب فى الاختبار على درجة كلية وإنما يكون التقرير وصفيا تفصيليا يحدد إنجازات الطالب وما استطاع أن يحققه أو يتقنه من معارف ومهارات وسلوكيات متضمنة فى نطاق المحتوى الذى يقيسه الاختبار الذى يكون عندئذ مرجعى المحك Criterion - Referenced وليس مرجعى المعيار Norm - Referenced.

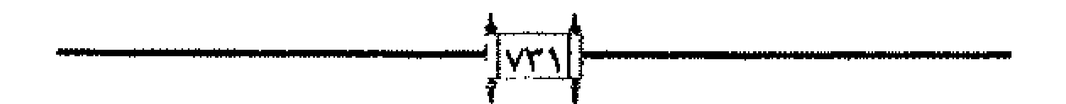
فالاختبارات مرجعية المحك - كما أوضحنا فى الفصل الثامن - تهتم بقياس مدى تحقق مجموعة متكاملة محددة من المهام وتُزود المتعلم بأدلة عن درجة تمكنه منها، كما توضح تقدمه أثناء دراسة مطلب تعليمى محدد وما يعترضه من صعوبات تعوق هذا التقدم، وبالتالى يمكنه أن يجرى تعديلا فى بيئة تعلمه ونوعية تعليمه مستفيدا فى ذلك من مقترحات معلمه وتعليقاته الإرشادية التى يوضحها على ورقة إجابة الاختبار. فالمعلم يسعى إلى تمكين جميع الطلاب من التحسن وذلك بتحديد متطلبات يمكنهم تحقيقها ويكون دور الطالب هو مشاركة المعلم فى تخطيط الأنشطة والعمل بجد لاستكمال هذه المتطلبات، وإبداء وجهة نظره فيما تعلمه فى إطار التعاون والمشاركة النشطة مع الأقران فى الفصل المدرسى الذى يشتمل على طلاب من ذوى قدرات متفاوته. حيث إن مستويات التحصيل لا يتحدد غالبا فى ضوئها تجميع الطلاب فى مجموعات.

فالقدرة يُنظر إليها على أنها مرتبطة ارتباطا وثيقا بنوعية التعليم وبيئته وطبيعة مهامّه وملاءمتها للفرد ودافعيته. وهذه المنهجية الجديدة فى قياس التحصيل تؤدى إلى

تكامل عمليتى التعليم والتقويم وهو هدف مهم تسعى إليه الأنظمة التربوية (Cronbach & Snow, 1977; Snow, 1980).

لذلك فإن الاختبارات مرجعية المحك أثرت تأثيرا ملحوظا فى تطوير مفهوم التحصيل حيث أصبح ينظر إلى قياس التحصيل كما سبق أن ذكرنا على أنه تقويم الكفايات Competency Assessment، أى قدرة المتعلم على القيام بمهام محددة ذات طابع تكاملى وظيفى بنجاح. فالتحصيل من هذا المنظور يعد بمثابة أداء Performance أى القيام بنشاط محدد. وهذا يتطلب أن يكون النطاق السلوكى لهذه الكفايات أو المهام مُعرفًا تعريفاً دقيقاً بحيث يصبح هذا النطاق هو المحك الذى تُنسب إليه درجة الفرد فى الاختبار. فتحديد النطاقات السلوكية يعد الخطوة الرئيسة فى بناء هذا النوع من الاختبارات. فقد اهتم علماء النفس فى الماضى ببناء نماذج سيكومترية تتعلق بدرجات الاختبارات وخصائصها الإحصائية، ولم يهتموا كثيرا ببناء نظريات يمكن الاسترشاد بها فى تحديد النطاقات السلوكية التى تقيسها الاختبارات التحصيلية. لذلك تبينت الحاجة الماسة إلى نظرية معرفية متكاملة Cognitive Theory يسترشد بها فى تعريف نطاق الأهداف التعليمية تعريفا يُمكّن من تكوين اختبارات محكة تسمح بالتفسير الدقيق للمطالب والمهام التى أتقنها الطالب، كما تبينت الحاجة إلى برمجيات للحاسوب يمكن باستخدامها ميكنة عملية تكوين المفردات الاختبارية Computerized Item Generation لتقليل عنصر الذاتية إلى أقل حد ممكن وقياس النطاق السلوكى الشامل للمحتوى التعليمى. ويُوجه خبراء القياس فى الدول المتطورة جانبا كبيرا من جهودهم لتطوير هذه التقنية من خلال أساليب متعددة أشرنا إليها فى الفصل الثامن وهى: برمجيات بناء المفردات Item grneration Programs ، والجمل الراسمة Mapping Sentences، وصيغ المفردات Item Forms، والتحويلات اللغوية للمفردات Item Linguistic Transformations.

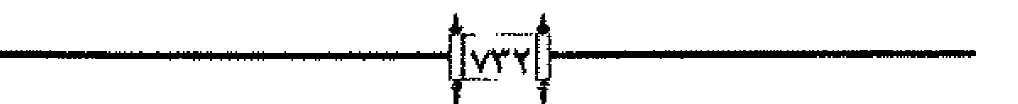
وتعتمد معظم هذه الأساليب على تخزين برامج خاصة للحاسوب ينتج عن تنفيذها تكوين عدد كبير من المفردات الاختبارية وفق قواعد ومواصفات محددة للنطاق السلوكى المراد قياسه. وعلى الرغم من التطورات التى حدثت فى هذا المجال، إلا أن هناك العديد من المشكلات المنهجية التى تعوق تنفيذ هذه الأساليب وتوسيع نطاق تطبيقاتها فى مختلف المجالات الدراسية. لذلك أصبحت هذه المشكلات مجالا خصبا لدراسات وبحوث القياس التربوى المعاصر.



(ثانيا) تقنيات تفريد قياس التحصيل باستخدام الاختبارات المفصَّلة Tailred testing :

إن التطورات العلمية والتكنولوجية والحركة البحثية الحادثة في مجال علم النفس المعرفي المعاصر ـ كما أوضحنا في الفصل السابق ـ سوف تؤدي إلى تطوير ملحوظ في أساليب قياس التحصيل الدراسي والعمليات الاختبارية وما تسهم به في تغيير الإطار التقليدي لبيئة التعلم وبخاصة عندما تتوافر الحاسبات الآلية المصغرة ويسهل اقتناؤها. فالتطورات التكنولوجية التي حدثت منذ أواخر الثمانينيات المتعلقة بتصميم شبكات الحاسبات، والتعقد المتزايد في وحدات هذه الحاسبات سواء استخدمت منفردة أو متصلة ببعضها كشبكة سوف تؤثر تأثيرا ملحوظا في البنى التحتية للمؤسسات التعليمية عندما يتوافر لكل من المعلم والمسئول الإداري والطالب حاسوب شخصي Personal Computer سواء في المدرسة أو المنزل بحيث ترتبط جميع هذه الحاسبات من خلال شبكة بحيث يمكن أن يتصل كل منهم بالآخر والمشاركة في استخدام البرامج والبيانات. وسوف يتيح هذا النظام التقني إمكانات متعددة جديدة لمدرسة المستقبل، حيث يؤدي إلى زيادة مصادر المعلومات نتيجة توسيع حلقة الاتصال بالعالم الخارجي سواء عن طريق القمر الصناعي، أو الوسائط الخاصة بالمقررات الدراسية ، أو الاتصال بين الطلاب في شتى الأماكن أو ربما بينهم وبين طلاب الدول الأخرى عن طريق شبكات الإنترنت . كما يؤدي إلى تجميع المعلومات من مصادر متعددة في إطار متكامل بحيث يشكل قاعدة بيانات ، وكذلك جعل الطلاب والكبار يتعلمون في المنزل كما لو كانوا في المدرسة (Sleeman & Brown, 1982).

وتبرز هذه التطورات التي سوف تحدث في البيئة التربوية وما سوف تقدمه التقنيات الجديدة من خبرات ثرية للطلاب العديد من التساؤلات حول الأنظمة الاختبارية فإذا تعلم الطالب الهندسة الفراغية مثلا من المنظور الديناميكي الذي يتميز بالشكل والحركة لموقع مجسم معقد يوضحه الحاسوب ، فكيف يمكن قياس تمكنه من هذه المعلومات باستخدام اختبار ورقة وقلم. إن هذا سوف يتطلب أدوات قياس من نوع جديد تحدد للطالب إنجازه المباشر في المهمة التي بين يديه من خلال استجابته لمثير أومثيرين بدلا من اختبار يستغرق ساعة أو ساعتين بحيث يمكنه التعرف مباشرة على مدى تمكنه من المهمة المعينة بما يتناسب وقدرته وسرعة تعلمه. ونظرا لأن المهام تختلف في طبيعتها فإن هذه الأدوات ينبغي أن تتنوع بتنوع المهام، مما يتطلب

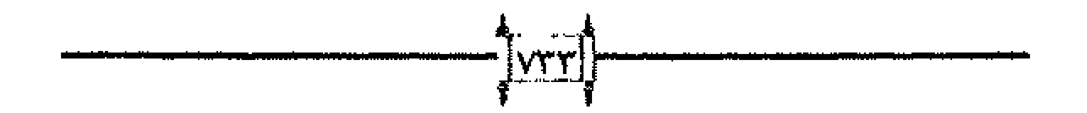


تقديم هذه المثيرات باستخدام الحاسوب. لذلك اتجهت بحوث القياس المعاصر فى الآونة الأخيرة إلى دراسة كيفية تفريد قياس التحصيل الدراسى وتطوير تقنياته مما سيؤدى إلى ابتكار أنواع جديدة من اختبارات التحصيل تختلف اختلافا ملحوظا عن الاختبارات المتعارف عليها.

فالقياس التقليدى للتحصيل الدراسى يتم بشكل جماعى، حيث يُطبق على جميع الطلاب الاختبار نفسه أو صيغ مكافئة له على الرغم من تباين مستويات قدراتهم. فلكى تكون للاختبارات الجماعية المقننة فائدة عملية فى إطار النظرية الكلاسيكية، يجب أن تشتمل على عدد كبير من المفردات المناسبة للأفراد من ذوى القدرة المتوسطة. ولكن يترتب على ذلك عدم فاعليتها فى قياس جميع مستويات قدرة الأفراد بالدقة نفسها. وربما يظن البعض أن إضافة بعض المفردات السهلة والمفردات الصعبة يعد حلا لهذه المشكلة، ولكن هذا ليس صحيحا. إذ إن المفردات السهلة ليس لها فائدة فى تقدير مستوى الأفراد من ذوى القدرة المرتفعة، والمفردات الصعبة لا تفيد فى تقدير مستوى الأفراد من ذوى القدرة المنخفضة. فتأثير عامل الملل فى الحالة الأولى، وعامل التخمين فى الحالة الثانية يزيد من الأخطاء العشوائية للقياس الناجمة عن إضافة هذه المفردات. كما أن اعتماد الخصائص السيكومترية لمثل هذه الاختبارات على خصائص عينة الأفراد المختبرين تجعل الموازنة بينهم ممكنة إذ اختُبر كل منهم باختبار مختلف يقيس القدرة نفسها.

وتفريد قياس التحصيل الدراسى يتطلب أن تكون صعوبة مفردات الاختبار مناسبة لمستوى قدرة الفرد الذى يختبر بها. وهذا يتطلب توافر مجموعة كبيرة من المفردات الاختبارية ذات خصائص سيكومترية معلومة بحيث يمكن سحب عينات منها تناسب المستويات المختلفة لقدرة الأفراد. وقد تمكن علماء القياس المعاصر من تطبيق أسس نظرية الاستجابة للمفردة الاختبارية Item Response Theory (IRT) التى أوضحناها بإيجاز فى الفصل السابق فى مجالات تطبيقية متنوعة عجزت النظرية الكلاسيكية للاختبارات عن مواجهتها. وسوف نوضح فيما يلى أهم هذه المجالات، وهى نظام الاختبارات المفصّلة Tailored Testing ، ونظام بنوك الأسئلة Item Banking.

فالطريقة المثلى لقياس قدرة الأفراد هى أن نزاوج بين صعوبة مفردات الاختبار ومستوى الدرجة الحقيقية True - Score level، أو القدرة التى يقيسها الاختبار، أو بعبارة أخرى أن نختبر الأفراد من ذوى القدرة المرتفعة بعدد أكبر من المفردات الصعبة، ونختبر ذوى القدرة المنخفضة بمفردات سهلة. وذلك لأنه إذا كانت مفردات الاختبار



صعبة جدا، أو سهلة جدا بالنسبة لفرد ما، فإننا لن نستفيد شيئا من هذه العملة الاختبارية سوى الحصول على بعض المعلومات عن الحد الأعلى، أو الحد الادنى لقدرة هذا الفرد. فإذا أجاب فرد ما إجابة خطأ عن مفردة متوسطة الصعوبة، لا يجوز لنا بعد ذلك مباشرة أن نختبره بمفردات صعبة جدا. لذلك يجب أن نختبر الفرد بمفردات اختبار تناسب مستوى قدرته، لكى نحصل على أكبر قدر من المعلومات عن هذا المستوى. وأفضل المفردات هى تلك التى يكون مقدار احتمال إجابة الفرد عنها إجابة صحيحة دون أن يلجأ إلى التخمين يساوى ٥٠ر. ، فليس من المعقول إذن أن نختبر مجموعة من الأفراد المختلفة فى قدرتهم بالاختبار نفسه. ولكن الأجدى أن نختبر هؤلاء الأفراد بمفردات مختلفة. ومن هنا نشأت الحاجة إلى الاختبارات المفصَّلة Tailored Testing ، وأحيانا تسمى الاختبارات التفريدية Individualized Tests ، أو الاختبارات المناسبة لقدرة الفرد Adaptive Testing. والمشكلتان المرتبطتان بهذا النوع من الاختبارات تتلخصان فيما يلى:

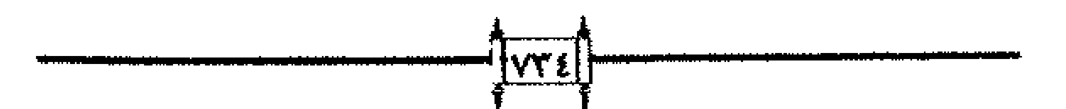
(١) اختيار مفردات تناسب كل فرد.

(٢) الحصول على تقدير لقدرة الأفراد على ميزان مشترك Common Scale ، على الرغم من اختلاف المفردات التى يختبر بها كل منهم.

وقد اقترح رايت Wright ودوجلاس Douglas طريقة لإعداد هذا النوع من الاختبارات، بحيث يتم تدريج صعوبة مفرداتها باستخدام نموذج راش. وتعتمد هذه الطريقة على إعطاء حرية للفرد فى انتقاء المفردات التى يود أن يُختبر بها. ثم يطبق عليه اختبار تتدرج فيه المفردات من حيث صعوبتها. ويُنصح الفرد بأن يبدأ بالمفردات الأكثر صعوبة والمفردات الأقل صعوبة، إلى أن يقرر بنفسه المفردات المناسبة له. وتقدر درجة الفرد فى الاختبار بعدد المفردات التى أجاب عنها إجابة صحيحة، مضافا إلى ذلك عدد المفردات التى قرر أنها سهلة جدا ولا تتحدى قدرته. لذلك تسمى هذه الطريقة «طريقة إعداد الاختبارات ذات المستوى المرن Flexilevel Testing».

ويرى لورد (Lord, 1977) أن إعداد نظام الاختبارات المفصلة، يحتاج إلى برنامج حاسوب يؤدى العمليات التالية:

(١) التنبؤ بكيفية إجابة الفرد عن المفردات التى لم يختبر بها بعد، بمعلومية إجابته عن المفردات التى اختُبِر بها بالفعل.



(٢) استخدام هذه المعلومات بطريقة مُجدية فى انتقاء مفردة الاختبار التى تطبق عليه فى المرحلة التالية.

(٣) تقدير درجة تمثل قدرة الفرد فى نهاية الفترة الاختبارية.

وقد حدثت تطورات كثيرة لفكرة الاختبارات المفصلة تناولت بصفة أساسية طبيعة الاستراتيجية التى يمكن ان تستخدم فى انتقاء المفردة التالية التى يُختبر بها الفرد. ومن بين هذه الاستراتيجيات:

(١) الاستراتيجية ذات المرحلتين Two - Stage Strategy.

(٢) الاستراتيجية متعددة المراحل Multi - Stage Strategy.

(٣) استراتيجية الاستعانة بالحاسوب Computer - Assisted Testing Strategy.

الإستراتيجية ذات المرحلتين:

اقترح كل من لورد (Lord, 1971)، وبتز، وويس (Betz & Weiss, 1974) استراتيجية تعتمد على تطبيق اختبار مبدئى Locator or Routing Test على الفرد، يحدد على أساسه المستوى الذى ينتقل إليه فى المرحلة الاختبارية التالية Second - Stage Test. وهذه تشتمل على مجموعة من الاختبارات التى تتوزع صعوبة مفرداتها على متصل القدرة. وتقدر قدرة الفرد باستخدام درجته فى كل من الاختبار المبدئى واختبار المرحلة التالية. وأحيانا تفيد درجة الاختبار المبدئى فى توجيه الفرد إلى المستوى الاختبارى التالى دون أن تؤخذ هذه الدرجة فى الاعتبار عند تقدير قدرته باستخدام درجته فى الاختبار التالى. وقد أجرى لورد Lord دراسات مستفيضة باستخدام المحاكاة عن طريق الحاسوب Computer Simulation Studies للبحث عن أفضل طرق تقدير القدرة فى هذه الاستراتيجية.

ووجد أن استخدام الدرجتين اللتين يحصل عليهما الفرد فى المرحلتين الاختباريتين أفضل من استخدام درجة الاختبار المبدئى فقط فى تقدير القدرة. ولكن هذا لا ينطبق فى حالة الأفراد من ذوى القدرة المتوسطة.

الإستراتيجية متعددة المراحل:

اعتمدت الاستراتيجية ذات المرحلتين على تفريعة واحدة One Branching، أى انتقال الفرد من الاختبار المبدئى إلى الاختبار التالى الذى يُستخدم فى قياس قدرته. أما

الاستراتيجية متعددة المراحل فهى تستطلب اتخاذ قرار عقب إجابة الفرد عن كل مفردة، لتوجيهه إلى المرحلة الاختبارية التالية. أى أن الاستراتيجية متعددة المراحل تعتمد على عدة تفريعات.

وتتضمن هذه الاستراتيجية نموذجين: أحدهما يسمى نموذج التفريع الثابت Fixed Branching Model، والآخر يسمى نموذج التفريع المتغير Variable Branching Model.

نموذج التفريع الثابت:

فى هذا النموذج يبدأ جميع الأفراد الإجابة عن مفردة متوسطة الصعوبة، وينتقلون بعدها إلى مجموعة من المفردات المتدرجة فى الصعوبة. وبعد الانتهاء من الإجابة عن مجموعة ثابتة من المفردات تقدر قدرة كل منهم إما باستخدام صعوبة المفردة (الافتراضية) التى كان من الممكن أن يُختبروا بها عقب المفردة التى ترتيبها (ن)، أو باستخدام متوسط صعوبة جميع المفردات بما فى ذلك المفردة الافتراضية التى ترتيبها (ن + ١) مع استبعاد المفردة الأولى من التقدير.

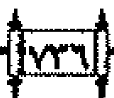
نموذج التفريع المتغير:

فى هذا النموذج يتم اختيار مفردة من بين مجموعة المفردات Items Pool فى كل مرحلة اختبارية يمر بها الفرد إلى أن نتأكد بدرجة كبيرة من الثقة من مستوى قدرته. ويمكن إعادة تقدير قدرة الفرد فى كل مرحلة إما باستخدام نظرية بييز Bayes Theory أو باستخدام طرق الأرجحية القصوى Maximum Likelihood Procedures. وقد عرض كل من أورى، (Urru , 1970)، وويس (Weiss,1974)، وجنسما (Jensema, 1974) وود (Wood , 1976)، ماكبرايد Mcbride لهذه الطرق بالتفصيل.

استراتيجية الإستعانة بالحاسوب:

تعد هذه الاستراتيجية أكثر الاستراتيجيات تعقيدا، إذا إنها تعتمد على الحاسوب فى اختيار المفردات التى يختبر بها الأفراد أثناء العملية الاختبارية Computer Assisted Testing.

فبعد أن يجيب فرد عن مفردة اختبار إجابة صحيحة، تقدم له مفردة أكثر صعوبة تختار طبقا لصيغة رياضية معينة. أما إذا أجاب عن المفردة إجابة خطأ، عندئذ تقدم له مفردة أقل صعوبة. وانتقاء مستوى صعوبة المفردات التالية يعتمد على «دالة درجية



«Step Function». وهذا يعنى أن صعوبة المفردة التى ترتيبها (ن + ١) تساوى صعوبة المفردة التى ترتيبها (ن) مضافا إليها أو مطروحا منها قيمة معينة تسمى «مدى الخطوة Step Size».

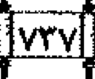
وتوجد طريقتان مختلفتان لتحديد مدى الخطوة: إحداهما تسمى «طريقة تقليص مدى الخطوة Shrinking- Step-Size Procedure »، والأخرى تسمى «طريقة مدى الخطوة الثابت Fixed-Step-Size Procedure».

وتبدأ الطريقة الأولى بخطوات مداها متسع، ثم يقل مدى الخطوة كلما اقترب مستوى صعوبة مفردة الاختبار من مستوى قدرة الفرد. ويمكن الحصول على تقديرات Estimates جيدة لقدرة الفرد باستخدام مجموعة معينة من المفردات. إلا أن هذا يتطلب وجود مجموعة كبيرة من المفردات Items Pool تنتقى من بينها المفردات المطلوبة. فإذا كان المطلوب من الفرد الإجابة عن (ن) من المفردات، يجب أن يكون عدد مفردات المجموعة $(٢)^{ن}$. فمثلا إذا كانت ن = ٦ يجب أن يكون لدينا ٦٤ مفردة، وإذا كانت ن = ١٠ يجب أن يكون لدينا ١٠٢٤ مفردة. وبالطبع يندر توافر هذا العدد الكبير من المفردات.

أما الطريقة الثانية فهى لا تتطلب إلا عددا قدره $\frac{ن(ن-١)}{٢}$ من المفردات فى المجموعة الكلية. فمثلا إذا كانت ن= ٥، يجب أن يكون لدينا ١٥ مفردة، وإذا كانت ن = ١٠، يجب أن يكون لدينا ٥٥ مفردة، وإذا كانت ن = ١٥، يجب أن يكون لدينا ١٢٠ مفردة. ولهذا فإن هذه الطريقة أقل كلفة وأكثر فاعلية من الطريقة الأولى.

وتقدير قدرة الفرد باستخدام هذه الطريقة يتطلب إيجاد متوسط صعوبة المفردات التى تم انتقاؤها بما فى ذلك صعوبة المفردة الأخيرة - أى المفردة التى ترتيبها (ن+١) - التى انتقاها الحاسوب، ولكن لم يتمكن الفرد من الإجابة عنها. وتقدر صعوبات هذه المفردات باستخدام أحد نماذج الاستجابة للمفردة الاختبارية (IRMs)، وبخاصة نموذج راش Rasch Model .

غير أن طرق تصميم وبناء هذه الاختبارات لا تزال موضع دراسات وبحوث سيكومترية متخصصة، حيث إن نظام تفريد قياس التحصيل وما يتطلبه من برامج حاسوب مبتكرة لا يزال فى مراحل تطوره الأولى. فهناك العديد من المشكلات الفنية المتعلقة بتنفيذه على نطاق واسع، ويسعى هؤلاء العلماء إلى التغلب عليها لكى يصبح مألوفا لدى المشتغلين بالقياس والمربين فى المستقبل القريب.



ومن أهم هذه المشكلات في الوقت الحاضر الكلفة الكبيرة نسبيا التي يتطلبها تنفيذ هذا النظام، وبخاصة ما يتعلق منها بفنيات تجهيز عتاد الحاسبات الآلية Hardware وإعداد البرامج المساعدة على تنفيذه Software.

ومن المشكلات المهمة أيضا التي تشغل فكر علماء القياس في هذا المجال كيفية تطوير الأسئلة التي تتطلب إجابات مفتوحة وتخزينها في الحاسوب من أجل التخفيف من سيطرة مفردات الاختيار من متعدد على الاختبارات التحصيلية في العقود الماضية، أو إعادة النظر في طريقة بناء وتصحيح هذه المفردات بحيث تقدم معلومات أكثر تفصيلا عن تحصيل الطلاب. ومن المتوقع مستقبلا تطوير أنظمة آلية ماهرة Expert systems ذات إمكانية تجهيز باللغات الطبيعية يمكنها تحليل المقطع النثرية وتقدير درجاتها تقديرا ذكيا وأقل كلفة مما هو عليه الآن.

وعندما يتم التغلب على هذه المشكلات فإن هذا النظام سوف يساعد على جعل عملية قياس التحصيل الدراسي أكثر مرونة وفاعلية ودقة بشرط توافر بنك للأسئلة تم تعيير مفرداته Calibrated Items وفقا لأحد نماذج الاستجابة للمفردة الاختبارية كما سنوضح بعد قليل.

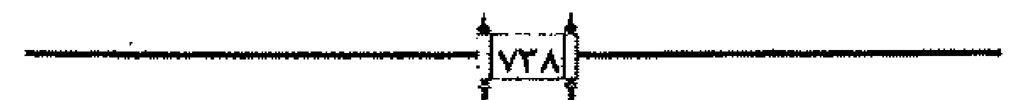
ومع هذا فإن هناك برامج حاسوب متوافرة في وقتنا الحاضر لعل أكثرها استخداما برنامج MicroCAT الذي يعد نظاما اختباريا يتم من خلال الحاسوب.

ويمكن للبرنامج أن يُكوّن بنوك أسئلة تشتمل على مفردات نصّية Text، وبيانية Graphics وصور متحركة Scanned Images، ويُشكّل من هذه البنوك اختبارات بمواصفات محددة ويُطبقها فرديا بما يتناسب مع مستوى قدرة الفرد الذي يُختبر به.

كما يجري البرنامج تحليل مفردات الاختبارات باستخدام النظرية الكلاسيكية ونظرية الاستجابة للمفردة الاختبارية. ويمكن الحصول على البرنامج كنظام متكامل أو كوحدات نسقية منفصلة مثل بنوك الأسئلة، وتفريد الاختبارات، و تحليل المفردات.

(ثالثا) بناء نظم بنوك الأسئلة Item Banking Systems:

إن نجاح علماء القياس التربوي والنفسي في تصميم تكنولوجية تساعد على بناء نظام الاختبارات المفصلة Tailored Testing ، أدى إلى زيادة الاهتمام ببناء نظم بنوك الأسئلة. والمقصود ببنك الأسئلة هو مجموعة من مفردات الاختبار يكون لها خصائص سيكومترية مميزة ومعلومة. ويجب أن تودع هذه المفردات في البنك بطريقة تيسر على المربين سحب مجموعات منها لتستخدم في بناء اختبارات تناسب الأغراض المختلفة



للقياس والتقويم التربوى. فتقويم التقدم الدراسى للطلاب يبدأ بالبحث عن أسلوب لقياس متغيرات القدرات المعرفية لدى الطلاب. وعلى الرغم من أن معظم نظم التقويم السائدة تعتمد فى قياس هذه المتغيرات على إيجاد العدد الكلى لمفردات الاختبار التى يجيب عنها كل طالب إجابة صحيحة، إلا أن هذا الأسلوب يقوم على أساس افتراضات يشوبها الخطأ كما سبق أن أوضحنا فى الفصل السابق.

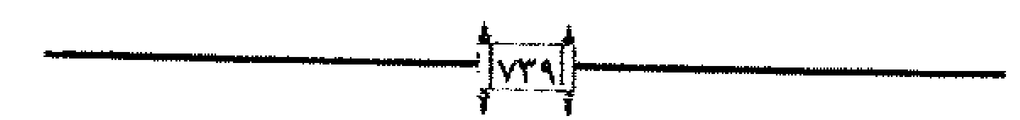
فمثلا يُفترض أنه يمكن جمع درجات جميع المفردات للحصول على الدرجة الكلية لمتغير واحد، وأنه يمكن الموازنة بين درجات الطلاب فى الاختبار، على الرغم من أن مفردات الاختبار التى تمثل المتغير المراد قياسه ربما لا تحمل المعنى نفسه بالنسبة لكل منهم، وكذلك يفترض أن درجات الاختبار تمثل القدرات تمثيلا خطيا.

ويعتمد استخدام الدرجات الخام فى تعريف «الميزان Scale» الذى يتم على أساسه الموازنة بين الطلاب فى عينة المفردات التى استخدمت فى القياس. وإذا حُولت الدرجات الخام إلى درجات موزونة، من مثل الدرجات المعيارية أو التائية، للتغلب على هذه المشكلة، تنشأ مشكلة أخرى وهى اعتماد هذا الميزان على عينة الطلاب المستخدمة فى إجراء هذا التحويل.

فالدرجات الخام تختلف عن كل من تدريج المفردات Items Calibration، وقياس قدرة الأفراد Person Measures، لأن هذه الدرجات تكون مقصورة على عينة الأفراد التى يتم اختبارها، فى حين أن تدريج صعوبة المفردات يجب أن يكون مستقلا عن خصائص العينة Sample-Freed. كما أن هذه الدرجات تقتصر على الاختبار الذى استمدت منه، فى حين أن قياس القدرة يجب أن يتصف بالعمومية بالنسبة إلى المتغير الذى تقيسه المفردات التى تم تدريج صعوبتها.

وقد رأينا فيما سبق أن نماذج الاستجابة للمفردة الاختبارية وبخاصة نموذج راش يفيد فى تحويل الدرجات الخام إلى تقديرات Estimates لمواقع الأفراد والمفردات على ميزان لوغاريتمى خطى بحيث تكون هذه التقديرات مستقلة عن كل من خصائص الاختبار، وخصائص عينة الأفراد. لذلك حاول علماء القياس النفسى والتربوى المعاصر بناء بنوك الأسئلة بالاستعانة بهذا النموذج.

فبعد الانتهاء من تعيير المفردات Items Calibration واختبار حُسن مطابقتها لنموذج راش، يصبح من السهل توحيد ميزان صعوبة المفردات Items Difficulties



Scale التى تم تقديرها من عينات مختلفة من الأفراد، وتودع هذه المجموعة من المفردات فى بنك الأسئلة. ويمكن عندئذ أن تسحب أى مجموعات جزئية من هذه المفردات لتكوين اختبارات، وإجراء قياسات متكافئة.

والطريقة التقليدية لجعل اختبارين متكافئين، هى أن نطبق كلا منهما على عينة من الأفراد، ونحدد الدرجات المتكافئة Equivalent Scores ، أى الدرجات التى تناظر رتبا مئينية واحدة. والطريقة الأخرى هى أن نطبق كلا من الاختبارين على عينة من الأفراد، ونستخدم نموذج راش فى الحصول على تدريج أو ميزان مشترك Cimmon Scale للمفردات. وهذا يعنى أن القياسات الناتجة عن استخدام أى من الاختبارين تكون متكافئة إحصائيا.

ويمكن ضم الاختبارين معا ليصبحا اختبارا واحدا، ثم يطبق على عينة الأفراد، واستخدام نتائج الاختبار ككل فى الحصول على التدريج المشترك. ولكن هذه الطريقة تصبح غير عملية إذا كنا بصدد بناء بنك للأسئلة يشتمل على عدد كبير من المفردات التى تقيس سمة كامنة أو قدرة معينة. فعندئذ يمكن تكوين اختبارات مختلفة من هذه المفردات، بحيث يشترك كل اختبارين منها فى عدد معين من المفردات يتراوح بين ١٠ ، ٢٠ مفردة. ثم نطبق كلا من هذه الاختبارات على عينة مختلفة من الأفراد، بحيث يُجيب كل فرد عن مفردات اختبار واحد فقط خلال شبكة من الوصلات بين المفردات Network of Links.

ولتوضيح ذلك، نفترض أن لدينا اختبارين يشتمل كل منهما على ٦٠ مفردة، وأردنا أن نجعل هذين الاختبارين متكافئين. فإحدى طرق إجراء ذلك هى أن نطبق كلا منهما على عينة تتكون من ١٢٠٠ فرد مثلا، ونحسب المئينيات المناظرة لكل درجة، ثم نحدد الدرجات التى تناظر رتبا مئينية واحدة.

ويمكن إجراء ذلك بطريقة أخرى تعتمد على بيانات أقل، وذلك بتكوين اختبار ثالث يشتمل على ٦٠مفردة مأخوذة من الاختبارين الأصليين (٣٠مفردة من كل). ثم نطبق كلا من هذه الاختبارات الثلاثة على عينة مختلفة تتكون كل منها من ٤٠٠فرد. وبذلك نكون قد اختبرنا كل فرد من أفراد العينة الكلية (١٢٠٠فرد) باختبار واحد، ومع هذا يمكن الحصول على تدريج مشترك لجميع المفردات (١٢٠مفردة) من خلال وصلتى المفردات Two Links (عدد مفردات كل منهما = ٣٠) التى تربط بين الاختبارات الثلاثة. ويمكن استخدام نموذج راش فى الحصول على تقديرين منفصلين

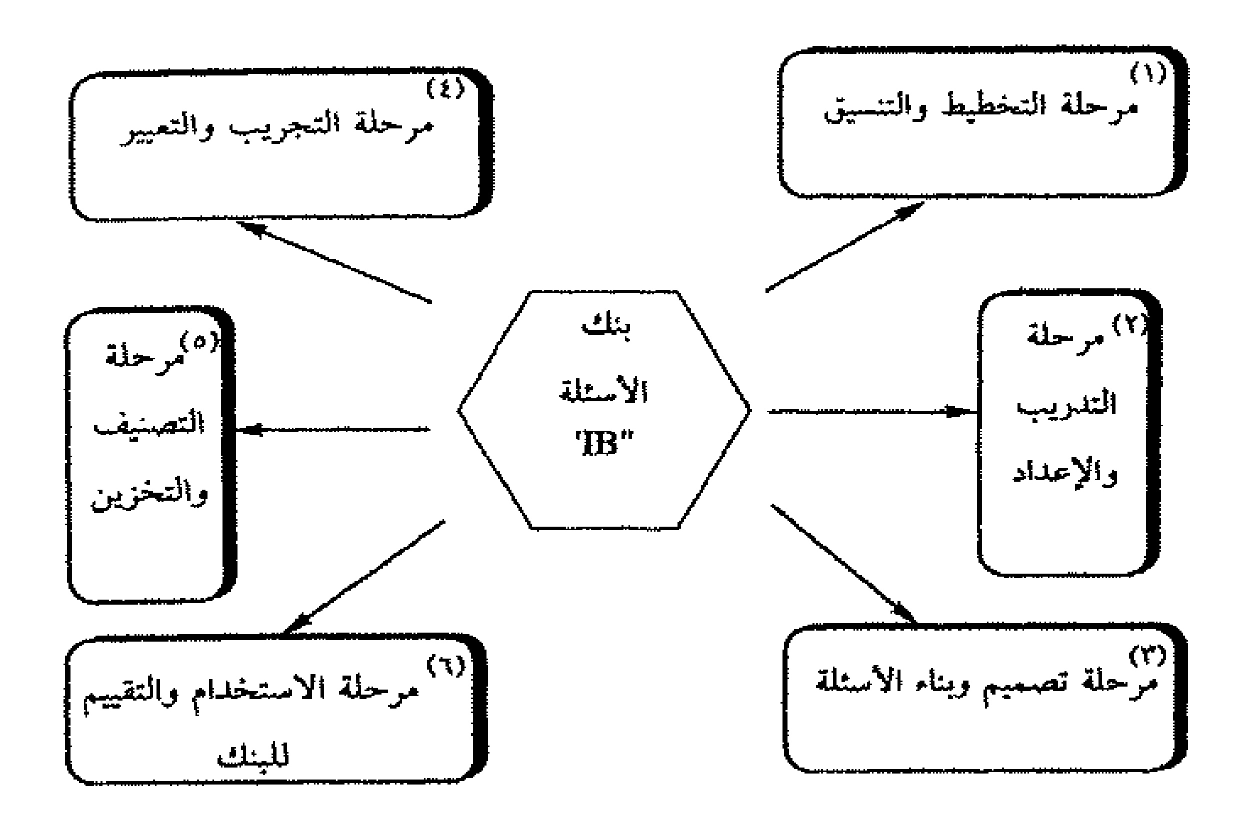
ومستقلين لصعوبة كل مفردة مشتركة بين الاختبارين الأصليين. وكل تقديرين يكونان متكافئين إحصائيا، ولكنهما يختلفان عن غيرهما بمقدار ثابت يسمى ثابت الانتقال Constant of Translation، ويكون لجميع المفردات التي يشتمل عليها كل من الاختبارين الأصليين المقدار الثابت نفسه. وهذا يمكّننا من تقدير ثابت الانتقال بسهولة وإضافته إلى جميع القيم التقديرية لصعوبة مفردات أحد الاختبارين، وبذلك يتم توحيد تدرايج ميزان صعوبة كل اختبارين من خلال المفردات المشتركة بينهما.

كما يمكن باستخدام نموذج راش تقدير حسن تكافؤ الاختبارين Goodness of Equating.

وبالطبع تنشأ بعض المشكلات أثناء إجراء هذا التحليل تتعلق بكيفية انتقاء المفردات المشتركة وعدد هذه المفردات، وكيفية التحليل إذا لم يكن الفرق بين القيم التقديرية لصعوبة المفردات المشتركة ثابتا، ومدى تأثير ذلك على تقدير قدرة الأفراد باستخدام الاختبارات المختلفة التي تشتمل على مجموعات متباينة من مفردات أو أسئلة البنك. هذا بالإضافة إلى ما يتطلبه إجراء هذا التحليل من عمليات إحصائية متعددة، وتوافر معدات حاسوب Hardware، وبرمجيات جاهزة مساعدة Software لتنظيم شبكات بنوك الأسئلة، وإدارة العملية الاختبارية بأسلوب ييسر توظيف هذه الشبكات توظيفا فاعلا.

مراحل إنشاء نظام بنك أسئلة مقترح:

نظراً لأهمية إنشاء نظام بنك أسئلة يسهم في تطوير أساليب قياس وتقويم التحصيل الدراسي. فسوف نقدم تصميمًا يقترحه المؤلف لبناء نظام بنكي للأسئلة متعدد الأبعاد والإجراءات. ويوضح الشكل التخطيطي (١٥ـ١) التالي المراحل التي يمر بها إنشاء هذا النظام البنكي المقترح:



شكل (١٥ـ١) مخطط مقترح لمراحل تأسيس وإنشاء نظام بنك الأسئلة

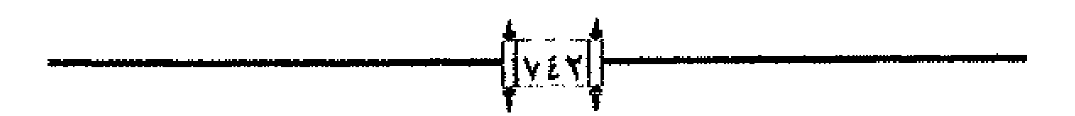
يتضح من شكل (١٥ـ١) أن نظام بنك الأسئلة المقترح يمر بست مراحل هي:

(١) مرحلة التخطيط والتنسيق:

فنظام بنك الأسئلة يتطلب وضع خطة متكاملة ومحكمة قبل البدء في تنفيذ المشروع، كما يتطلب تخطيطا وتنسيقا لتحديد المسئوليات والمهام والإسهامات الواجبة في المشروع. كذلك يتطلب تهيئة المسئولين الإداريين والمربين والمتعلمين والجمهور بهذه النقلة النوعية في نظم وأساليب التقويم والإمتحانات من خلال اللقاءات والمشاركات.

(٢) مرحلة التدريب والإعداد:

فنظرا إلى أن المشروع يعتمد بدرجة أساسية على توافر كفايات على درجة عالية من التخصص والخبرة فإنه ينبغي وضع خطة تنظيمية لتدريب الأطر الفنية التي ستوكل إليها مهام تحديد الأهداف والكفايات التعليمية وصياغتها بطريقة تسمح بقياس نواتج



التعلم، وكتابة مختلف أنواع الأسئلة والمفردات الاختبارية التي تقيس الأهداف التي يتم تحديدها. وينبغي التأكد من امتلاكهم القدرة والمهارة الفائقة في صياغة هذه الأسئلة والمفردات. وبخاصة تلك التي تقيس المستويات المعرفية العليا سواء الموضوعية منها أو المقالية.

(٣) مرحلة تصميم الأسئلة وبنائها:

فالأسئلة والمفردات الاختبارية هي المادة الخام التي تستخدم في بناء الاختبارات، لذلك فإن مرحلة تصميم هذه الأسئلة وبنائها ومراجعتها مراجعة متأنية يعد من أدق المراحل وأهمها على الإطلاق. لذلك ينبغي أن يقوم بهذه المرحلة خبراء متمرسون وعلى أعلى مستوى من الكفاية في بناء مختلف أنواع الأسئلة والمفردات الاختبارية وبخاصة في المستويات المعرفية والمهارية العليا.

(٤) مرحلة التجريب والتعيير:

فأهم ما يميز بنك الأسئلة احتواؤه على رصيد متجدد من أسئلة ومفردات اختبارية ذات خصائص قياسية معلومة. فالانتهاء من كتابة هذه والأسئلة في ضوء أهداف المحتوى الدراسي ومراجعتها لا يعني أنها أصبحت صالحة للإيداع في البنك، إذ ينبغي أن تخضع لعمليات مستمرة من التجريب الميداني على عينات مناسبة من الطلاب، واستخدام الأساليب السيكومترية والإحصائية لتحليل البيانات المتعلقة بكل سؤال أو مفردة لتقييم فاعليتها، من حيث مدى ملاءمتها للمجتمع الذي تمثله هذه العينات من المتعلمين، وقدرتها على التمييز بين المستويات التحصيلية المتباينة، وذلك وفق نموذج سيكومتري مناسب.

فنظام بنك الأسئلة لا يعد نظاما اختباريا أو نظام قياس ما لم يحقق خاصيتي المرونة والاتساق. ونقصد بالمرونة إمكانية تشكيل اختبارات ذات خصائص محددة وتناسب أغراضا معينة من مخزون البنك من الأسئلة والمفردات الاختبارية. أما الاتساق فنعني به توحيد نظام تعيير أو تدريج جميع مفردات البنك بما يسمح بتكوين اختبارات متكافئة يسهل تفسير درجاتها من اختبار إلى آخر، كما يسمح بالموازنة بين أداء فردين في اختبارين مختلفين في درجة صعوبة مفردات كل منهما ولكنهما يقيسان نطاقا سلوكيا واحدا.

ولعل استخدام أحد نماذج الاستجابة للمفردة الاختبارية، وبخاصة نموذج راش يمكن أن يحقق ذلك إذا أحسن استخدامه.

(٥) مرحلة التصنيف والتخزين:

فتنظيم بنك الأسئلة يعد من الأمور المهمة، إذ من الصعب أن يحقق البنك الفائدة المرجوة منه دون تنظيم دقيق وإدارة محكمة. فأساليب الإيداع والسحب من رصيد الأسئلة والمفردات الاختبارية، والتطوير المستمر لمخزون البنك، وعمليات الإضافة والحذف من رصيده، وإجراءات مراقبة هذه العمليات ومتابعتها والإشراف عليها والتحكم فى حركتها تتطلب جميعها إجراءات تنظيمية دقيقة، ومحكات وضوابط فنية، كما تتطلب أن تكون الهيئة الإدارية المسئولة عن بنك الأسئلة ذات كفاءة عالية.

(٦) مرحلة الاستخدام والتقييم للبنك:

فمحك فاعلية نظام بنك الأسئلة يتحدد فى ضوء الاستخدام الميدانى الفعلى للأسئلة والمفردات الاختبارية المودعة فيه. ومما لا شك فيه أن نتائج الاستخدام وما يترتب عليه تكون بمثابة تغذية مرتدة تعين فى تقييم فاعليته وإعادة النظر فى مكوناته وأسلوب تنظيمه بصفة مستمرة.

إذ يمكن تشكيل اختبارات كاملة بمواصفات محددة من رصيد البنك يمكن على أساسها تقويم تحصيل الطلاب فى المجالات الدراسة المختلفة، والإفادة من التطبيق الميدانى فى المراجعة المستمرة لأسئلة البنك ومفرداته سواء بالإضافة أو الحذف أو إعادة التعبير، بل وربما يؤدى إلى إعادة النظر فى المنهج المدرسى وفقا لما تسفر عنه نتائج الاستخدام.

كذلك ينبغى تنظيم الخدمات الاختبارية التى يمكن أن يقدمها البنك للمعلمين والطلاب، مثل الاختبارات التدريبية أثناء العام الدراسى، وتصحيح الاختبارات الموضوعية، وبخاصة مفردات الاختيار من متعدد، والصواب أو الخطأ، وإعداد دليل متجدد كل عام يوضح للمعلمين والمعنيين بنظام البنك مواصفات ما يشتمل عليه من أسئلة ومفردات اختيارية ومجالاتها. وبذلك يسهم هذا النظام البنكى للأسئلة فى تطوير عمليات قياس وتقويم التحصيل الدراسى وجعلها أكثر مرونة وفاعلية ودقة، وكذلك جعلها أكثر تقنينا مما هى عليه فى إطار الممارسات الحالية، كما يسهم فى تزويد المربين ببيانات تتعلق بالتقدم الدراسى للطلاب فى وقت قصير. وبذلك يكون البنك مكونة رئيسة من مكونات المنظومة التعليمية، ويسهم إسهاما فاعلا فى تطويرها، كما تصبح عمليتا التعليم والتقويم عملية واحدة متكاملة، وهو ما يسعى إليه خبراء القياس المعاصر.

البنك الدولي للأسئلة وتقويم التحصيل التربوي من منظور دولي:

يعد مشروع البنك الدولي للأسئلة International Item Bank من التطورات المهمة الأخرى في مجال تقويم التحصيل التربوي على الصعيد الدولي. وهو أحد المشروعات الدولية التي بدأتها الجمعية الدولية لقياس التحصيل The International Association for the Evaluation of Educational Achievement(IEA) بالتعاون مع مركز دراسات التقويم بجامعة كاليفورنيا Center for the Study of Evaluation (CSE). وقد اشترك في هذا المشروع الذي بدأ تنفيذه عام ١٩٨٠ عشرون دولة من الدول الأعضاء في هذه الجمعية الدولية.

واهتم هذا المشروع ببناء مركز دولي منبع وشبكة بنوك أسئلة Item Banks Network تساعد في تطوير أساليب وعمليات تقويم التحصيل التربوي في الدول الأعضاء. ورأى المسئولون في هذه الجمعية أن بناء بنك دولي للأسئلة متصل ببنوك الأسئلة الخاصة بكل دولة سوف يؤدي إلى ابتكار نظم تصنيفية للأسئلة ولخصائصها السيكومترية على موازين دولية موحدة International Common Scales بحيث يسهل على جميع الدول الأعضاء استخدامها، ويتميز بالمرونة الكافية بحيث يتفق والأنظمة التعليمية في هذه الدول، كما يساعد في الحصول على بيانات مقارنة عن جوانب تعليمية معينة في الدول المختلفة.

وقد قامت هذه الجمعية أيضا بسلسلة من الدراسات المقارنة لتطوير مقاييس معيارية للتحصيل التربوي تصف مستوى التحصيل في مختلف هذه الدول، وتحدد العوامل التي تفسر الاختلافات عبر الثقافية في التحصيل. فأُجريت مثلا دراسة مقارنة في الرياضيات في ٢٠ دولة، وجمعت أيضا بيانات تتعلق بالطالب والمعلم وخصائص المدرسة. وكذلك أجريت دراسة مقارنة في ست مواد دراسية هي: العلوم، والآداب، وفهم المقروء، واللغتين الإنجليزية والفرنسية كلغتين ثانيتين، والتربية الاجتماعية في ٢٠ دولة بما في ذلك بعض الدول النامية. ومن المشكلات التي واجهت مثل هذه الدراسات صعوبة بناء اختبارات تحصيل تناسب جميع الدول على الرغم من سهولة ذلك إلى حد ما في الرياضيات، وكذلك مشكلة تباين المناهج وما تركز عليه من دولة إلى أخرى، ومشكلة معاينة الطلاب Students Sampling. وقد أوضحت النتائج أن العوامل التي تُسهم في رفع مستوى التحصيل تتباين بتباين الدول. ومن بين هذه العوامل: مستوى تعليم الوالدين، والمستوى الاجتماعي الاقتصادي، ومصادر التعلم المتنوعة المتوافرة. وهذه الدراسات الدولية تفيد بلا شك في تقديم معلومات ثرية

يمكن استخدامها فى ترشيد القرارات المتعلقة بتطوير التعليم دون الحاجة إلى الاستناد إلى درجات كل طالب على حدة.

رابعا) التقييم التوثيقي الشامل للتحصيل والكفايات

: Portfolio Assessment

يعد هذا التوجُّه فى تقييم تحصيل الطلاب وكفاياتهم من الاتجاهات المعاصرة التى برزت فى الميدان التربوى فى الآونة الأخيرة، ويختلف اختلافا جوهريا عن غيره من التوجهات التى أوضحناها فى الجزء السابق. ويرى مؤيدوه أن معظم أساليب القياس والتقويم السائدة تنظر إلى الفرد نظرة جزئية تفتقر إلى التكامل. فالاختبارات التى تعتمد على مفردات الاختيار من متعدد أو غيرها من المفردات تركز على معرفة الحقائق وحل مشكلات اصطناعية. لذلك يتبنى هذا التوجه مدخلاً جديداً فى التقييم يُطلق عليه التقييم البديل Alternative Assessment، أو التقييم التوثيقى الشامل Portfolio Assessment، أو التقييم الواقعى Authentic Assessment.

وهذا النوع من التقييم يهدف لقياس إمكانات عقلية عليا ويركز على عمليات تعلم مهمة يمكن تنميتها فى إطار العمل المدرسى وخارجه ومتابعة تطورها، ويعتمد فى ذلك على صحيفة وثائق تتعلق بالطالب Portfolio. ويرى بعض المربين أن هذه الصحيفة الوثائقية تشتمل على جميع الأنشطة أو الأعمال المدرسية للطالب طوال العام الدراسى، وتُستخدم فى تسجيل هذه الأعمال. ويرى البعض الآخر أن هذه الصحيفة تشتمل على عينة ممثلة لأعمال الطالب توضح إنجازاته لكى يطلع عليها ولى أمره ويتعرف نوعية هذه الإنجازات. كذلك يمكن أن تشتمل هذه الصحيفة على مجموعة منتقاة من الأعمال لغرض معين، مثل عينة من كتابات الطالب أو مهاراته فى حل المشكلات، أو مجموعة منتقاة من أفضل أعمال الطالب لعرضها على أقرانه أو أسرته أو أفراد مجتمعه المحلى.

ولعل هذا يماثل ما يعرضه الفنانون والمصوِّرون من عينات جيدة لأعمالهم، غير أن عينات أعمال الطلاب تختلف عن ذلك من حيث الغرض منها وتنوع محتواها وصيغتها وفقا للمجال الدراسى أو المهارة المرجوة. وقد استخدمت الصحائف Protfolios بدرجة أكبر فى مجالات المهارات اللغوية، مثل الكتابة والقراءة، والمهارات الفنية الرفيعة، وكذلك فى الرياضيات، والعلوم الإنسانية.

أما التقييم التوثيقي الشامل Portfolio Assessment فإنه يهدف للتحقق من جودة ونوعية تحصيل الطالب واكتسابه المعارف والمهارات والقدرات والتوجهات العقلية المحددة في عبارات تشير إلى مستويات الاكتساب، وذلك من خلال فحص عينات من أعمال الطالب التي يختارها بإرشاد من معلمه، ويتم هذا الفحص بواسطة فرد يختلف عن المعلم Audience . ويسترشد في تقييمه بمحكات متفق عليها ومرتبطة بواقع المجتمع. وغالبا تركز عملية التقييم على شكل الصحائف الوثائقية للطالب ومحتواها. ولكن استخدام هذه الصحائف في تقييم التحصيل يمكن أن Protfolio يفيد فائدة كبيرة في تعميق فكر المربين فيما يتعلق بالتدريس والتعلم.

مكونات الصحائف الوثائقية:

تشتمل الصحائف الوثائقية للطالب Portfolios على ثلاث مكونات (Gredier, 1995; Storms et a1., 1996) هي:

(١) سيرة أعمال الفرد Biographies of Works: وهي توثيق لتطور أو إنتاج مشروع رئيس يقوم به الطالب ويمثل أفكار الطالب في المشروع المختار. وسيرة أعمال الفرد تعبر عن مدى عمق جهوده في مجال دراسي معين.

(٢) أعمال متنوعة للطالب Range of Works: وتتضمن الأنشطة والأعمال التي ينجزها الطالب، وتوضح مدى عمق إنتاجه. فمثلا يمكن أن تشتمل صحيفة وثائقية لطالب في مجال المهارات اللغوية على مقالات كتبها، أو شعرًا نظمه، أو حوارًا مسجلا داخل الصف، أو غير ذلك.

(٣) انعكاسات الطالب Student Reflections: فخلال المدة الدراسية يطلب المعلم من الطالب وصف الجوانب المميزة لأعماله وإنتاجه، وتوضيح مدى تقدمه ونموه، والأهداف التي تتطلب التحقيق. كذلك ربما يطلب منه القيام بدور الناقد المتخصص أو المؤرخ الشخصي. وتفيد هذه المكونة في إثراء تعلم الطالب ونموه.

صيغ الصحائف الوثائقية:

تتنوع صيغ الصحائف الوثائقية Portfolio Formats التي أشارت إليها الأدبيات في هذا المجال (Gredler, 1996; Scanlan & Heiden, 1996; Valencia & Calfee, 1991)

ولعل أهمها ما يلي:



(١) **الصيغة المثالية** Ideal Format : وتهدف لمعاونة الطالب ليصبح قادرا على تقييم سيرة دراسته تقييما واعيا مستبصرا، وذلك لإثراء نوعية تعلمه وتعرف نموه. وتشمل هذه الصيغة سيرة أعمال الطالب، وأعمال متنوعة، وانعكاسات الطالب التى تكون مؤشرا لقدرته على تحليل وتقييم أعماله.

(٢) **صيغة العرض** Showcase Format: وتهدف لعرض عينات من أعمال الطالب على الآباء وغيرهم من أفراد المجتمع المحلى. لذلك تشتمل هذه الصيغة على مجموعة من أفضل إنتاج الطالب الذى يختاره بنفسه. فانعكاسات الطالب واختياراته تعد عنصراً أهم من استخدام مقاييس مقننة.

(٣) **صيغة التوثيق** Documentation Format : وتهدف لتقديم سجل منظم لإنتاج الطالب ونتائج التقويم الكمى والكيفى لأعماله. وتشتمل على سجل منظم مستمر لتقدم الطالب إلى جانب بعض انعكاسات الطالب وتقويم المعلم، ونتائج الملاحظات، وقوائم المراجعة، والسجلات القصصية أو الوقائعية، واختبارات الأداء.

(٤) **صيغة التقويم** Evaluation Format : وتهدف لتقويم تقرير مقنن عن تحصيل الطالب وإنتاجه لولى أمره وللمسئولين فى المدرسة. وتشتمل على مجموعة من أعمال الطالب التى يختارها المعلم أو المسئولين فى المدرسة أو الإدارة التعليمية وفق محكات محددة مسبقا.

(٥) **صيغة صفية** Class Format: وتهدف لإعلام الآباء والمسئولين فى المدرسة بأحكام المعلم عن إنجازات كل طالب. وتشتمل على بطاقات تلخص أداء جميع الطلاب فى ضوء أهداف المنهج، وملاحظات المعلم وتعليقاته حول أداء كل طالب، وخطة المعلم التعليمية طوال العام الدراسى مع بعض اقتراحاته المتعلقة بتطوير العمل.

ومن هذا يتضح أن الصحائف الوثائقية Portfolios تختلف باختلاف الهدف منها الذى يتراوح بين إثراء نوعية عملية التعلم، وتقديم معلومات مقننة للمسئولين فى المدرسة أو الإدارة التعليمية عن إنجازات الطلاب فى ضوء أهداف المقررات الدراسية. ففى الحالة الأولى يكون التركيز على نمو كل طالب وتقييم الطالب لأدائه وإنجازاته تقييما ذاتيا وتحليله لمدخلاته، أما فى الحالة الثانية فيكون التركيز على كتابة تقارير تتضمن معلومات مقننة تتعلق بمجموعات الطلاب وذلك لتقديمها للمسئولين.

وعلى الرغم من اختلاف أهداف هذه الصحائف إلا أنها تقدم مؤشرات متعددة عن تحصيل الطالب، كما يمكن أن تفيد في تقييم الأداء الذي يتطلب مدى متسعا من الوسائط والأنماط، وكذلك تقييم أعمال الطالب التي يستغرق إنجازها مدة زمنية معينة. وربما يصعب إجراء ذلك باستخدام الاختبارات التحصيلية التقليدية.

ويعتمد نجاح التقييم التوثيقي الشامل Portfolio Assessment على عوامل متعددة؛ من أهمها تطوير مهارات المعلمين وتغيير اتجاهاتهم نحو أساليب التعليم التي تركز على الحفظ والتلقين إلى أساليب تركز على النمو الشامل للطالب وإعطائه الثقة في أهمية تحليله لأدائه وتقويمه الذاتي وتوثيق إنجازاته وتتبع نموه بنفسه، ومشاركة المعلمين والآباء والطلاب في لقاءات دورية لمناقشة الصحائف الوثائقية للطلاب، وتصميم إجراءات التقييم التوثيقي الشامل من حيث كيفية جمع المعلومات والبيانات، وتحليل محتوى الصحائف، وتحديد محكات تقييمها والحكم على نوعيتها، والإفادة من نتائجها. وبذلك تصبح هذه الصحائف الوثائقية Portfolios أداة فاعلة في التعليم والتقويم.

بعض مشكلات تقييم التحصيل باستخدام الصحائف الوثائقية

إن تحديد ما يجب تقييمه، أي أهداف نظام التقييم التوثيقي الشامل Portfolio Assessment يعد أحد المشكلات التي تواجه مستخدمي هذا النظام. فقد لوحظ من خلال البرامج التي تنفذها بعض الدول المتطورة في هذا المجال عدم اتفاق المسئولين حول ما يجب أو يمكن تقييمه في الصحائف الوثائقية Portfolios . لذلك اعتمد تقييم بعض هذه البرامج على الأهداف والمستويات القومية المحددة، ومراشد المناهج، وتوجيهات خبراء المواد الدراسية، واستنباط عدد محدود من العناصر الثرية من هذه المصادر المتعددة للاسترشاد بها في التقييم، وعدم الاقتصار على وضع قائمة بالمهارات المرجوة.

ونظرا لأن هناك تعريفات متعددة للصحائف الوثائقية Portfolios، فإن هناك طرقا متعددة لتقييمها وفقا للظروف المحلية وما يناسب القائمين بالتقييم. ففي بعض الأنظمة تقدر درجة لكل جزء من الصحيفة الوثائقية للطالب، ويوجد متوسط الدرجات للتوصل إلى تقدير عام. ويهتم بعض آخر بإصدار أحكام شاملة في مجالات أداء معينة في ضوء محكات محددة في دليل معلم. وعلى الرغم من اختلاف هذه الأنظمة، إلا أنها تؤكد أهمية الاستناد إلى محكات أداء متفق عليها مجتمعيا توضح مستويات الأداء المختلفة، وأحيانا تحدد دليل تقدير درجات ممثلة في خمسة أو ستة مستويات. والمشكلة الثانية تتعلق بدور المعلم في نظام الصحائف الوثائقية. فالمعلم يقوم بتصميم

ما يجب أن يتناوله الطالب في صحيفته، وإرشاد الطلاب أثناء قيامهم بتكوين هذه الصحائف، وتنظيم الجدول الزمني لهذه الصحائف في إطار اليوم المدرسي، وكذلك المشاركة في بناء محكات لتقييم الصحائف نظرا لأنه أقدر من غيره على تحديد الأداء المناسب والأداء غير المناسب، وتعرف الظروف الأكثر ملاءمة لتنفيذ هذا النظام داخل غرفة الصف، ومراجعة نتائج التقييم. وهذا يتطلب بالطبع تنمية مهنية معينة للمعلم للقيام بهذه الأدوار الجديدة.

والمشكلة الثالثة تتعلق بأساليب تقييم صيغ الصحائف الوثائقية المتنوعة التي أشرنا إليها. فالصيغ التوثيقية والتقويمية والصفية لا تهتم بانعكاسات الطالب أو نمو مهارات تحليله لأدائه، وإنما تهتم بإنجازاته وتحصيله. وأحيانا تستخدم الصيغة التقويمية في تقديم بيانات تتعلق بنوعية تحصيل الطالب والإسهام في تطوير عملية التعليم داخل الصف المدرسي. غير أن هناك تعارضًا إلى حد ما بين أهداف قياس نوعية التحصيل، وتطوير عملية التعليم. فتطوير أداء البرنامج لكي يحقق هدفا معينا ربما يقلل من فاعلية أدائه في تحقيق هدف آخر، وهذا ينعكس على بنية الصحائف الوثائقية.

والمشكلة الرابعة تتعلق بصدق وثبات درجات أو تقديرات الصحائف. فالتساؤل الأساسي المتعلق بصدق الصحائف التي توثق إنجازات الطالب هو التكوينات الفرضية التي يمثلها إنتاج الطلاب. ولكن نظرا لاتساع تعريفات نطاق محتوى المجالات الدراسية مما يؤدي إلى تفسيرات متعددة فإنه يصعب الاستناد إلى هذه التعريفات في التحقق من صدق تقييم هذه الصحائف.

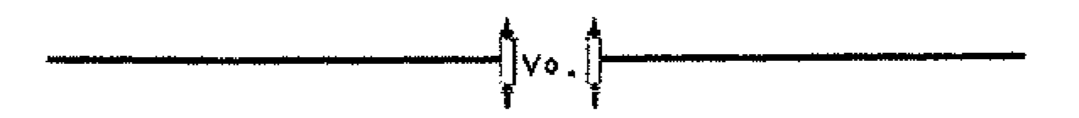
لذلك اتجه بعض الباحثين إلى إيجاد الارتباط بين درجات الصحائف ودرجات الاختبارات المعتادة في المجال نفسه، غير أن القيم الناتجة كانت منخفضة.

وفيما يتعلق بالثبات فقد استخدم بعض الباحثين ثبات تقديرات المحكمين للصحائف، ولكن القيم الناتجة كانت منخفضة أيضا في بعض المجالات. وربما يرجع ذلك إلى عدم وضوح المصطلحات في نظام تقدير الصحائف Scoring Rubrics، أو تعقد هذا النظام، أو عدم كفاية تدريب القائمين بالتقدير، أو تباين المهام في الصفوف المختلفة، أو اختلاف الظروف التي يتم العمل في ضوئها.

وفيما يلي نقدم مثالا لبعض التساؤلات في مقابلة لمناقشة الصحيفة الوثائقية للطالب التي تتضمن مجموعة من أعماله وأنشطته:

(١) ما العمل الذي تعتز به أكثر من غيره؟

(٢) ما العمل الذي تود مراجعته بدرجة أكبر، أو تحتاج مزيدًا من الوقت لتحسينه؟ ولماذا؟



(٣) ما العمل الذى لست راضيا عنه؟ ولماذا؟

(٤) إذا أتيحت لك فرصة تنفيذ أيٍّ من هذه الأعمال مرة أخرى، ما التغييرات التى يمكنك إجراءها عليه؟

(٥) أى هذه الأعمال كان تنفيذه أصعب؟ ولماذا؟

(٦) أى هذه الأعمال كان أيسر تنفيذا؟ ولماذا؟

(٧) أى هذه الأعمال تعلمت منه بدرجة أكبر من غيره؟ ولماذا؟

تساؤلات تقدم موشرات عن علاقة تعليم الطالب بعمله:

(١) أى هذه الأعمال قمت بتنفيذه فى جماعة؟

(٢) كيف يختلف العمل فى جماعة عن العمل وحدك؟

(٣) أى هذه الأعمال يتطلب منك تخطيطا وتنظيما أكثر؟

(٤) كيف استفدت من هذا التخطيط والتنظيم؟

(٥) ماذا تعلمت من هذا التخطيط والتنظيم؟

(٦) هل تتضمن صحيفتك الوثائقية مشروعا استغرق أسبوعين أو أكثر؟

(٧) ما الاستراتيجية التى اتبعتها لإدارة وقتك وتنظيمه أثناء عملك فى المشروع؟

ومن الجدير بالذكر أن تنفيذ متطلبات نظام التقييم التوثيقى الشامل فى الصف المدرسى Classroom - Based Portfolio Assessment يتطلب تطوير تقنيات الحاسوب فى المدارس، وابتكار صيغ تجديدية لهذا النظام يمكن استدعاؤها بسهولة وتكون محدودة المساحة.

وعموما،فإن هذا النظام المستحدث يعد أداة فاعلة فى توثيق النمو التحصيلى للطالب من خلال وسائط متعددة تشمل: قوائم المراجعة، والملاحظة، والسجل القصصى، وموازين التقدير، وكتابات الطلاب ودراساتهم، والمحاكاة باستخدام الحاسوب، والعروض الشفوية والسمعية والبصرية، والصور، والرسوم، والقصص، والتسجيلات، والمقابلات، والمؤتمرات، وغير ذلك. وثراء وتنوع هذه الأساليب والوسائط يمكننا من تقويم إنجازات الطالب تقويما شموليا Holistic Evaluation فى مضمون أكثر واقعية من مضمون التقويم السائد فى وقتنا الحاضر الذى يقتصر على الاختبارات التقليدية.

خلاصة

تناولنا فى هذا الفصل بعض أهم التوجهات التطبيقية للقياس والتقويم التربوى والنفسى استنادا إلى التطورات المعاصرة التى أوضحناها فى الفصل السابق. فقد أسهمت نظرية الاستجابة للمفردة الاختبارية إسهاما مثمرا فى مجالات تطبيقية متنوعة عجزت عن مواجهتها النظرية الكلاسيكية للاختبارات، وقد أوضحنا بعض المجالات المتعلقة بتطوير أساليب قياس وتقويم التحصيل الدراسى والكفايات وهى: استخدام الاختبارات مرجعية المحك فى تقويم التحصيل، وتقنيات تفريد قياس التحصيل باستخدام الاختبارات المفصلة وبناء نظم بنوك الأسئلة. وقدمنا تصميما مقترحا لإنشاء نظام بنكى للأسئلة متعدد الأبعاد والإجراءات استنادا إلى نموذج راش. كما أشرنا إلى مشروعات البنك الدولى للأسئلة، وتقويم التحصيل من منظور دولى. ونظرا لأن هناك توجهًا معاصراً نحو توسيع نطاق مفهوم التحصيل وارتباطه بالكفايات والأداء المتميز فقد تناولنا مجالا مستحدثا أصبح يطبق فى عديد من الدول المتطورة يتعلق بالتقييم التوثيقى الشامل للتحصيل والكفايات Portfolio Assessment ، وأوضحنا مكونات الصحائف الوثائقية Portfolios وصيغها، وبعض المشكلات المتعلقة بتصميمها وتنفيذها وتقييمها، والإفادة من تقنيات الحاسوب فى التغلب على هذه المشكلات.

ومن الجدير بالذكر أن تعدد وتنوع التطورات المعاصرة فى مختلف مجالات القياس والتقويم التربوى والنفسى التى عرضناها فى هذا الكتاب يدل على ثراء هذا العلم وارتباطه بالحاجات المتزايدة للمجتمعات المتقدمة فى هذا القرن، وتأثره تأثرا ملحوظًا بالتطورات التقنية التى أسهمت إسهاما كبيرا فى تقدم حركة التنظير السيكومترى والإديومترى، وتطبيقاتها المتعددة فى المجالات التربوية والنفسية والاجتماعية.



1 - Adomo, T. , Frenkel- Bruswik, E. , Levinson, D. and Sanford, R. The Authoritarian Personality. N.Y. : Harper & Row, 1950.

2 - Aiken, L. Readings in Educational and Psychological Testing. Boston: Allyn and Bacon, Inc. , 1973.

3 - Aiken, L. Psychological Testing and Assessment. (5th Ed.) , N.Y. : Allyn and Bacon, 1986.

4 - Airasian, P. & Madaus, G. " Linking Testing and Instruction: Policy Issues" **Journal of Educational Measurement,** 1983,20,2, 103-118.

5 - Allam, S. Minimum Measurement Competencies for Classroom Teachers. N.J. : Educational Testing Service , 1980.

6 - Allport, G. Personality: A Psychological Interpretation. N.Y. : Holt, Rinehart & Winston, 1937.

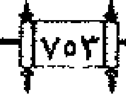
7 - Anastasi, A. Testing Problems in Perspective. N.J. : ETS, 1965.

8 - Anastasi, A. Psychological Testing (4th ed.) . N.Y. : Macmillan Publishers, 1976.

9 - Anastasi, A. Fields of Applied Psychology. N.Y. :McGraw - Hill, 1979.

10 - Anderson, S. , Ball, S. and Murphy, R. Encyclopedia of Educational Evaluation . San Francisco: Jossey-Bass Publishers, 1975.

11 - Apter , M . The Computer Simulation of Behavior . London: Hutchinson & Co. Publishers, 1970.



12 - Arndt, W. Theories of Personality. N.Y. : Macmillan Publishing Co. Inc. , 1974.

13 - Arter , J . The Portfolios for Assessment and Instruction . ERIC Digest, 1995.

14 - Assessment Systems Corporation. Computerized Testing Products Catalog. MN: ASC, 1992.

15 - Ayala, R. ; Dodd, B. and Koch, W. "A Simulation and Comparison of Flexilevel and Bayesian Computerized Adaptive Testing" **Journal of Educational Measurement,** 1995, 27, - 239.

16 - Babbie, E. Survey Research Methods. Calif. : Wadsworth Publishing Company , Inc., 1973.

17 - Barnette, W. Readings in Psychological Tests and Measurements (Revised ed.). Ill. : The Dorsey Press, 1968.

18 - Barry, N. & Shannon, D. "Portfolios in Teacher Education: A Matter of Perspective" .**The Educational Forum**, 1997, 61 , 320-328.

19 - Bauernfeind, R. Building a School Testing Program. Ill. :Educational Testing Service, 1963.

20 - Beaton, A. & Zwick, R. "Overview of the National Assessment of Educational Progress". **Journal of Educational Statistics,** 1992, 17,2,95-109.

21 - Beatty , W. Improving Educational Assessment and Inventory Measures of Affective Behavior. Washington: NEA, 1969.

22 - Berdie, D. & Anderson, J. Questionnaire: Design and Use . N.J.: Scarecrow, Metuchen, 1974.

23 - Berk , R . Consumers' Guide to Criterion- Referenced Test Item Statistics. NCME Publications, 1978, 9, 1.

24 - Bertrand, A. & Cebula, J. Tests, Measurement, and Evaluation: A Developmental Approach. Calif. : Addison-Wesley, 1980.

25 - Birenbaum, M. & Dochy, F. Alternativs in Assessment of Achievement, :Learning Processes and Prior Knowledge. The Netherlands: Kuwer Academic Publishers, 1995.

26 - Birrell, J. & Ross, S. "Standardized Testing and Portfolio Assessment: Rethinking the Debate". **Reading Research and Instraction,** 1996, 35,385-396.

27 - Blalock , H. Measurement in the Social Sciences: Theories and Strategies. Ill. : Aldine Publishing Company, 1974.

28 - Bloom, B. ; Madaus, G. and Hastings, T. Evaluation to Improve Learning. N.Y. : McGraw-Hill, 1981.

2٩ - Borgatta, E. & Lambert, W. Handbook of Personality Theory and Research. Ill. : Rand McNally & Co. , 1968.

30 - Bormuth, J. On the Theory of Achievement Test Items. Chicago, IL.: University of ChicagO Press, 1970.

31 - Boud, D. Enhancing Learning Through Self Assessment. London: Kagan Page, 1995.

32 - Brennan, R. Some Applications of Generalizability Theory to the Dependability of Domain - Referenced Tests. A Paper Presented at the Annual Johns Hopkins University National Symposium on CRM, 1978.

33 - Brennan , R . Handbook for GAPID : A Computer Program for Generalizability Analysis with Single Facet Design. ACT Technical Report. Iowa: ACT, 1979.

34 - Brennan , R . Elements of Generalizability Theory . Iowa: ACT Publications, 1983.

35 - Brown, F. Principles of Educational and Psychological Testing (2nd ed.) . N.Y. : Holt, Rinehart and Winston, 1976.

36 - Burns, E. The Development, Use and Abuse of Educational Tests. Ill.: Charles Tomas, 1979.

37 - Burns, R. The Self Concept: Theory, Measurement and Behavior. N.Y.: Longman, 1979.

38 - Butcher, H. & Lomax, D. Readings in Human Intelligence. N.Y. : Harper & Row Publishers, Inc. , 1972.

39 - Byrne, D. An Introduction to Personality. N.Y. : Prentice. Hall, Inc. , 1966.

40 - Carlson , R . & Locklin , R . Item Response Theory: Comparing BILOG and MicroCAT Calibration for a Mathematics Ability Test. MI. : ERIC, 1995.

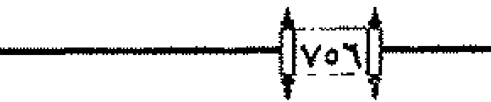
41 - Cartwright, C. & Cartwright, P. Developing Observation Skills. N.Y.: McGraw-Hill, 1974.

42 - Cattell, R. Handbook of Modern Personality Theory. Chicago: Aldine, 1970.

43 - Cattell, R. ; Eber, H. ; and Tatsuoka, M. Handbook for the Sixteen Personality Factor Questionnaire . Ill . : Institute for Personality and Ability Testing, 1970.

44 - Chase, C. Measurement for Educational Evaluation (2nd ed.) . Calif.: Addison-Wesley, 1978.

45 - Cherryholmes, C. "Construct Validity and Discourses of Research". **American Journal of Education,** 1988, 96,3, 421-457.



46 - Coffman, W. Frontiers of Educational Measurement and Information System. Boston: Houghton Mifflin Co., 1973.

47 - Cohen, G. The Psychology of Cognition . N.Y. : Academic Press, 1977.

48 - Collins, H. Educational Measurement and Evaluation: A Worktext. Ill.: Scott, Foresman, 1976.

49 - Cooke, D. ; Craven, H. & Clarke, G. Basic Statistical Computing. Scotland: Thomson Ltd., 1982.

50 - Coombs, C. A Theory of Data. N.Y. : John Wiley, Inc., 1964.

51 - Couch, A. & Keniston, K. "Response Sets and Social Desirability". **Journal of Abnormal and Social Psychology,** 1961, 62,165-169.

52 - Cox, R. & Vargas, J. A Comparison of Item Selection Techniques for Norm-Referenced and Criterion- Referenced Tests . Paper Presented at the Annual Meeting of the NCME . Chicago, 1966.

53 - Crombag, H. & Gruijter, N. Contemporary Issues in Educational Testing. Paris: Mouton, 1974.

54 - Cronbach, L. Essentials of Psychological Testing (3 rd ed.) N.Y. : Harper & Row, Publishers, 1970.

55 - Cronbach, L. Toward Reform of Program Evaluation. Calif.: Jossey-Bass Publishers, 1980.

56 - Cronbach, L. Five Perspectives On Validity Argument. In H. Wainer (Ed.) , Test Validity. N.Y.: Erlbaum, 1988.

57 - Cronbach, L. Construct Validity After Thirty Years.In R. Linn (Ed.), Intelligence : Measurement, Theory and Public Policy. Urbana: University of Illinois Press, 1989.



58 - Cronbach, L. & Drenth, P. Mental Tests and Cultural Adaptation. Netherlands: Mouton Publishers, 1972.

59 - Cronbach, L . & Glaser , R . Psychological Tests and Personnal Decisions (2nd ed.) Ill. : University of Illinois Press, 1965.

60 - Cronbach, L. & Meehl, P. "Construct Validity in Psychological Tests". **Psychological Bulletin,** 1955, 52.

61 - Cronbach, L. ; Glaser, R . ; Nanda, H. and Rajaratnam, N . The Dependability of Behavioral Measurement : Theory of Generalizability for Scores and Profiles. N.Y.: Wiley, 1972.

62 - Cropley , A . "Creativity and Intelligence" . **British Journal of Educational Psychology,** 1966, 36, 259 - 266.

63 - Crowne, D. The Experimental Study of Personality. N.Y. : Lawrence Erlbaum Publishers, 1979.

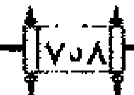
64 - Crowne , D . & Marlow , D . The Approval Motive: Studies in Evaluative Dependence. N.Y. : John Wiley, Inc., 1964.

65 - Dahistrom, W. ; Welsh, G. and Dahistrom, L. An MMPI Handbook: Clinical Interpretation (A Revised ed.) . Minn. : University of Minnesota Press, 1972.

66 - D'Angelo, E. The Teaching of Critical Thinking. Amestrdam : Gruner, 1971.

67 - Dann, P. ; Irvine S. and Collins, J. Advances in Computer - Based Human Assessment. MA. : Kluwer Academic Publishers, 1991.

68 - Davis, I. Competency Based Learning: Technology, Management, and Design . N.Y. : McGraw- Hill, 1973.



69 - Dockrell, W. On Intelligence. London: Methuen & Co. Ltd. , 1970.

70 - Doner, D. ; Karn , M. and David, A. Students React to Portfolio Assessment. **Contemporary Education,** 1998, 69,3, 159, 165.

71 - Dong , T. & Sarti, L. Design and Production of Multimedia and Simulation - Based Learning Material. The Netherlands: Kuwer Academic Publishers, 1994.

72 - Dunn. Rankin. P. Scaling Methods. N.Y. : Erlbaum, Hillsdale, 1983.

73 - Ebel, R. Essentials of Educational Measurement, (3 rd ed.) N.J. : Engleswood Cliffs, 1979.

74 - Edwards, A. Techniques of Attitude Scale Construction. N.Y.: Appleton - Century- Crofts, Inc. 1957.

75 - Edwards, A. The Social Desirability Assessment and Research. N.Y. : Dayden, 1957.

76 - Edwards, A. The Measurement of Personality Traits By Scales and Inventories. N.Y. : Holt , Rinehart and Winston, 1970.

77 - - Edwards, A. & Walker, J. "Social Desirability and Agreement Response Set " **Journal of Abnormal and Social Psychology,** 1961, 62, 180 , 183.

78 - ETS The Redesign of Testing For the 21st Century. N.J.: ETS, 1985.

79 - Eysenck, H. The Structure of Human Personality (3 rd ed.) , London: Methuen & Co. Ltd., 1970.

80 - Eysenck, H. The Measurement of Intelligence. Lancaster: MTP, 1973.

81 - Eysenck, H. The Measurement of Personality. Lancaster: MTP, 1976.

82 - Eysenck, H. The Structure and Measurement of Intelligence . N.Y.: Springer - Verlage, 1979.

83 - Fahmlow, H. & Woodley, K. Objective - Based Testing: A Guide to Effective Test Development . N . J . : Educational Technology, 1969.

84 - Feder, P. The Complete Guide to Taking Tests. N.J. : Prentice- Hall, 1979.

85 - Ferguson, L. Personality Measurement (2 nd ed.) . N.Y. : McGraw - Hill, 1971.

86 - Fincher, J. Human Intelligence. N.Y. : Putnam's Sons, 1976.

87 - Fine, B. The Stranglehold of the I.Q. N.Y. : Doubleday & Co. , Inc. , 1975.

88 - Fishbin, M. Readings in Attitude Theory and Measurement. N.Y.: John Wiley, 1967.

89 - Fiske, D. Measuring the Concept of Personality. Chicago: Aldine Publishing Company, 1971.

90 - Forgus, R. & Shulman. B. Personality: A Cognitive View. N.J. : Prentice- Hall, Inc., 1979.

91 - Frahm , R . & Covington , J. What's Happening in Minimum Competency Testing. **Phi Delta Kappa,** 1979.

92 - Frediksen, N; Mislevy, R. and Bejar, I. Test Theory for a New Generation of Tests. N.J. : Erlbaum, 1993.

93 - Freeman, J. ; Butcher, H., and Christie, T. Creativity: A Selective Review of Research. London: Society for Rescarch into Higher Education Ltd. ; 1968.



94 - Fricke, B. Opinion, Attitude and Interest: Survey Handbook, A Guide to Personality and Interest Measurement. MI. : OAIS Testing Program, 1965.

95 - Gable, R. & Wolf, M. Instrument Development in the Affective Domain: Measuring Attitudes and Values in Corporates and School Settings. The Netherlands: Kuwer Academic Publishing, 1994.

96 - Gekoski, N. Psychologial Testing (2 nd ed.) . Ill: Charles Thomas, 1964.

97 - Georgi , D. & Crowe , J . "Digital Portfolios : A Confluence of Portfolio Assessment and Technology". **Teacher Education Quarterly,** 1998, 25, 73 - 84.

98 - Getzels, J. & Jakson, P. Creativity and Intelligence, N.Y. :Wiley, 1962.

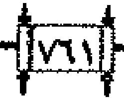
99 - Goldman, L. Using Tests in Counselling (2 nd ed.). N.Y. : Meredith Corporation, 1971.

100 - Gorow , F. Better Classroom Testing. Penn. : Chandler Publishing Co. , 1966.

101 - Gorth. W. Comprehensive Achievement Monitoring: A Criterion - Referenced Evaluation Technology, 1975.

102 - Gredler, M. "Implications of Portfolio Assessment for Program Evaluation" **Studies in Educational Evaluation,** 1995, 21, 431 - 437.

103 - Green , B. Adaptive Testing by Computer: New Directions for Testing and Measurement. No. 17, Edited by R. Ekstrom. San Francisco: Jossey- Bass, 1983.



104 - Gregg, L. Knowledge and Cognition . N.Y. : John Wiley, 1974.

105 - Gronlund, R. Readings in Measurement and Evaluation. N.Y. : The Macmillan Company, 1968.

106 - Gronlund, N. Preparing Criterion-Referenced Tests for Classroom Instruction. N.Y. : Macmillan, 1973.

107 - Gronlund, N. Measurement and Evaluation in Teaching (4 th ed.). N.Y. : Macmillan , 1981.

108 - Gruijter , D . & Kamp . L . Advances in Psychological and Educational Measurement. N.Y.: Wiley & Sons, 1976.

109 - Guilford, J. Personality. N.Y. : McGraw- Hill, 1959.

110 - Guilford, J. The Nature of Intellignce. N.Y. : McGraw - Hill, 1967.

111 - Guilford, J. & Hoepfner, R. The Analysis of Intelligence. N.Y. : McGraw-Hill , 1971.

112 - Guion, R. Personnal Testing . N.Y. : McGraw-Hill ,1965.

113 - Gulliksen, H. Theory of Mental Tests. N. Y. : Wiley, 1950.

114 - Guslkey, T. Implementing Mastery Learning. Calif.: Wadsworth Publishing Company, 1985.

115 - Hambleton, R. & Gruijter, D. "Applications of Item Response Models to Criterion- Referenced Test Item Selection". **Journal of Educational Measurement,** 1983, 20 , 4 , 355-367.

116 - Hambleton , R . & Zaal , J. Advances in Educational and Psychological Testing: Theory and Applications. MA. : Kluwer Academic Publishers, 1991.

117 - Haney, W. "Testing Reasonsing and Reasoning About Testing" . **Review of Educational Research,** 1984, 54, 4 , 597-654.



118 - Harman , H . Modern Factor Analysis , (2nd ed.). Chicago: University of Chicago Press, 1967.

119 - Harris, C. & Wilcox, R. Achievement Test Items: Methods of Study. Calif. : CSE, 1977.

120 - Henerson, M. ; Morris, L.; and Tayler, C. How to Measure Attitudes. Calif. : Sage Publications, 1978.

121 - Hewitt, G. Shoud We Use Pupils Writing Protfolios to Grade Schools? They Guide Improvement. **American Teacher,** 1978, 83 , 2 , P. 4.

122 - Heywood, J. Assessment in Higher Education. N.Y.: John Wiley, 1977.

123 - Hills, J. Measurement and Evaluation in the Classroom . Ohio : A Bell & Howell Company, 1976.

124 - Hively, W. Domain- Referenced Testing . N.J. : Englewood Cliffs, 1974.

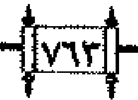
125 - Holtzman , W . Computer-Assisted Instruction , Testing and Guidance N.Y.: Harper & Row , Publishers, 1970.

126 - Hopkins , K . & Stanley , J. Educational and Psychological Measurement and Evaluation (6 th ed .) . N . J . : Prentice-Hall Inc., 1981.

127 - Horrocks, J. & Schoonover, T. Measurement for Teachers. Ohio: Charles Merrill Publishing Company, 1968.

128 - Houts , P . The Myth of Measurability: IQ Tests. N.Y.: Hart Publishing Company, 1977.

129 - Hunt, E. & Pellegrino, J. "Using Interactive Computing to Expand Intelligence Testing : A Critique and Prospects " . **Intelligence,** 1982, 9 , 3 , 207 - 230.



130 - Izar , J . Assessing Learning Achievement. Paris: UNESCO, Document 60 , 1990.

131 - Jackson, D. & Messick, S. Problems in Human Assessment. N.Y. : McGraw-Hill, 1967.

132 - Jensema, C. "An Application of Latent Trait Mental Test Theory". **British Journal of Mathematical and statistical Psychology.** 1974, 27.

133 - Jensen, A. Bias in Mental Testing. N.Y.: Free Press, 1980.

134 - Kantowitz , B . Human Information Processing : Tutorials in Performance and Cognition. N.Y. : John Wiley, 1974.

135 - Kaplan, R. Psychological Testing : Principles, Applications, and Issues. Calif. : Brooks, 1983.

136 - Kaufman, A. Intelligence Testing with the WISC-R. N.Y.: Wiley, 1979.

137 - Keats, J. An Introduction to Quantitative Psychology. N.Y. : John Wiley, 1971.

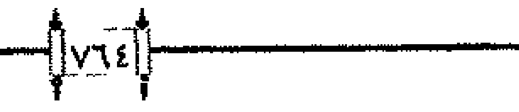
138 -Keeves, J. Educational Research, Methodology and Measurement: An International Handbook. N.Y. : Pergamon, 1997.

139 - Kelly, E. Assessment of Human Characteristics. Calif. : Brooks / Cole Publishers. , 1967.

140 - Kelly. G. The Psychology of Personal Constructs. N.Y. : Noron & Co. , Inc., 1955, 2 Vols.

141 - Kimlig, L. & Jarvik, L. Genetics and Intelligence: **A Review of Science**, 1963 , 142, 1477-1479 .

142 - Klahr, D. Cognition and Instruction. N.Y.: Lawerence Erlbaum Publishers, 1976.



143 - Kleinmuntz, B. Personality Measurement : An Introduction. Ill. : the Dorsey Press, 1967.

144 - Kline, P. New Approaches in Psychological Measurement. N.Y. : John Wiley, 1973.

145 - Kline, P. Psychological Testing : The Measurement of Intelligence, Ability and Personality. N.Y. : Crane Russak, 1976.

146 - Kluchnikoff, B. Education and Learning for the 21st Century: A Priority Agenda. Paris: UNESCO, 1992.

147 - Koch, W. & Dodd. B. Operational Characteristics of Adaptive Testing Procedure using Partial Credit Scoring.Paper Presented at the Annual Meeting of the AERA. San Francisco, 1986.

148 - Labaw, P. Advanced Questionnaire Design. Mass. : Abt, 1980.

149 - Lachar, D. The MMPI: Clinical Assessment and Automated Interpretation. Calif. : WPS, 1974.

150 - Lancaster, O. Effective Teaching and Learning . N.Y. : Gordon and Breach, 1974.

151 - Lazarsfeld, P. The Logical and Mathematical Foundation of Latent Structure Analysis. In S.A. Stouffer et al., Measurement and Prediction. Princeton: Princeton University Press, 1950.

152 - Lawler , J . , IQ, Heritability and Racism. N.Y. : International Publishers, 1978.

153 - Lesgold, A. ; Pellegrino, J. ; Fokkema, S. and Glaser, R. Cognitive Psychology and Instruction. N.Y. : Plenum Press, 1977.

154 - Lewis, G. Assessment in Education. N.Y.: Wiley, 1975.

155 - Lien, A. Measurement and Evaluation of Learning . Iowa: Brown Company Publishers, 1961.

156 - Lindeman, P. & Merenda, P. Educational Measurement, (2 nd ed.). Ill. : Scott Foresman, 1979.

157 - Linden, K. & Linden, J. Modern Mental Measurement: A Historical Perspective. MA. : Houghton Mifflin, 1968.

158 - Linn, R. Educational Measurement (3 rd ed.). Washington, D. C. : NCME, 1989.

159 - Lord, F. & Novick, M. Statistical Theories of Mental Test Scores. Mass. : Addison Wisley, 1968.

160 - Lord, F. Applications of Item Response Theory to Practical Testing Problems. N.J. : Hillsdale, 1980.

161 - Macintosh, H. Techniques and Problems of Assessment. London: Edward Arnold, 1974.

162 - Madaus G. & Stuffelbeam, D. Educational Evaluation: Classic Works of Ralph Tylor . MA . : Kluwer Academic Publishers, 1991.

163 - Magnusson, D. Test Theory. Pal Alto: Addison-Wesley, 1966.

164 - Matarazzo , J. Wechsler's Measurement and Appraisal of Adult Intelligence. Baltimore: The Williams & Wilkins Co. , 1972.

165 - McArthur, D. Alternative Approaches to the Assessment of Achievement . MA. : Klwuer Academic Publishers, 1991.

166 - McCaughey, S. Using Process-folios with High Ability Students in Visual Arts. **Gifted Child Today Magazine**, 1997 , 20, 40-42.



167 - McKinney , M . Preservice Teachers' Electronic Portfolios : Investigating Technology, Self-Assessment, and Reflection. **Teacher Education Quarterly**, 1998, 25, 85-103.

168 - McReynolds, P. Advances in Psychological Assessment (Volume1) Calif. : Science and Behavior Books, Inc., 1968.

169 - Meehl , P . Clinical Versus Statistical Prediction. Minneapolis: University of Minnesota Press, 1954.

170 - Meeker, M. The Structure of Intellect: Its Interpretations and Uses. Ohio: Merrill Publishing Company, 1964.

171 - Mehrens , W . & Ebel , R . Principles of Educational and Psychological Measurement. Ill. : Rand McNally& Co. , 1967.

172 - Mehrens, W. & Lehman, I. Standaradized Tests in Education, (2 nd ed.). N.Y. : Holt, Rinehart and Winston, 1975.

173 - Mehrens , W . & Lehmann , L . Measurement and Evaluation in Education and Psychology (4 th ed.). Fort Worth: Holt, Rinehart & Winston, 1991.

174 - Micklo , S . Math Portfolios in the Primary Grades. **Childhood Education**, 1997, 73, 194-199.

175 - Miller, K. Psychological Testing in Personnel Assessment . N.Y. : John Wiley, 1975.

176 - Miller, D. Handbook of Research Design and Social Measurement (3 rd ed.). N.Y. : David Mckay Co. , Inc. , 1977.

177 - Millman , J . & Arter , J . Issues in Item Banking. **Journal of Educational Measurement,** 1984, 21, 4, 315-335.

178 - Mills, E. Augmented Title: Submission of Portfolios on a Computer Disk. **International Journal of Instructional Media,** 1997, 24, 23-29.

179 - Morris, L. & Tayler, C. How to Measure Achievement. Calif. : Sage Publications, 1978.

180 - Moss , P . Shifting Conceptions of Validity in Educational Measurement: Implications for Performance Assessment. **Review of Educational Research,** 1992, 62, 3, 229-258.

181 - National Computer Sytems (NCS). MICRO TEST SCORE II, User's Guide for the Sentre Plus System. MN.:NCS, 1986.

182 - Nettles , M. & Nettles, A. Equity and Excellence in Educational Testing and Assessment . The Netherlands: Kuwer Academic Publishers, 1995.

183 - Nitko, A. Educational Tests and Measurement: An Introduction. N.Y. : Harcourt Brace Jovanovich, Inc. , 1983.

184 - Noll, V. & Scannell, D. Introduction to Educational Measurement (3 rd ed.). N.Y. : Houghton Mifflin Co. , 1972.

185 - Nunnally, J. Psychometric Theory . N.Y. : McGraw-Hill, 1978.

186 - Oakland, T. & Hambleton, R. International Perspectives on Academic Assessment.The Netherlands: Kuwer Academic Publishers, 1995.

187 - O' Neil, H. Learning Strategies. N.Y. : Academic Press, 1978.

188 - Osgood, C. ; Suci, G.; and Tannenbaum, P. The Measurement of Meaning. Ill. : University of Illinois Press, 1957.

189 - Osterlind, S. Constructing Test Items . MA. : Kluwer Academic Publishers, 1991.

190 - Passow, A. ; Noah , H. ; Eckstein, M. The National Case Study: An Empirical Comparative Study of Twenty- One Educational Systems. N.Y. : John Wiley, 1976.

191 - Passow, H . The Gifted and the Talented : Their Education and Development. Ill. : The University of Chicago Press, 1979.

192 - Patterson, C . The Wechsler- Bellevue Scales : A Guide for Counselors. Ill. : Charles Thomas Publishing, 1953.

193 - Payne, D. The Assessment of Learning: Cognitive and Affective. MA. : D. C. Heath, 1974.

194 - Payne, D. Designing Educational Project and Program Evaluation. The Netherlands: Kuwer Academic publishers, 1994.

195 - Payne, D . & McMorris , R . Evaluation and Measurement Contributions to Theory and Practice. N.J. : General Learning Press, 1975.

196 - Pervin , L . Personality : Theory , Assessment , and Research (2 nd ed.). N.Y. : John Wiley, 1975.

197 - Popham, J. Evaluation in Education. Calif. : McCutchan Publishing Corporation, 1974.

198 - Popham, J. Criterion-Referenced Measurement . N.J. ; Prentice-Hall, 1978.

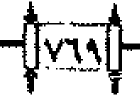
199 - Popham. J. Modern Educational Evaluation. N.J. : Prentice- Hall, 1978.

200 - Rasch, G. Probabilistic Models for Some intelligence and Attainment Tests. University of Chicage, 1980.

201 - Resnick, L. The Nature of Intelligence. N.J. : Lawrence Erlbaum, 1976.

202 - Richard, R. & Andrey , H. Fundamentals of Behavioral Statistics (5th ed.). N.Y. : Addison - Wesley Publishing Company, 1984.

203 - Riley. M ; Riley, J. and Toby, J. Scale Analysis: Applications, Theory and Procedures.N.J.:Rutgers University Press, 1954.



204 - Rohwer , W . ; Ammon , P. and Cramer, P. Understanding Intellectual Development. Ill. : the Dryden Press, 1974.

205 - Roid, G. & Haladyna, T. Technology of Test Item Writing. N.Y. : Academic Press, 1982.

206 - Rorrer, L. The Great Response Style Myth. **Psychological Bulletin,** 1965, 63, 129-156.

207 - Roskam, E. & Vandenwollenberg M. Recent Developments in Item Response Theory and Applications in the Netherland . Paper presented at the Annual Meeting of the AERA, 1985.

208 - Rotter, L. "Generalized expectancies for Internal Versus External Control of Reinforcement". **Psychological Monographs,** 1966, 80, 1.

209 - Rushby , N. Computers in the Teaching Process. N.Y. : John Wiley, 1979.

210 - Safrit , M. & Wood T. Introduction to Measurement in Physical Education. N.Y.: Mosby, 1995.

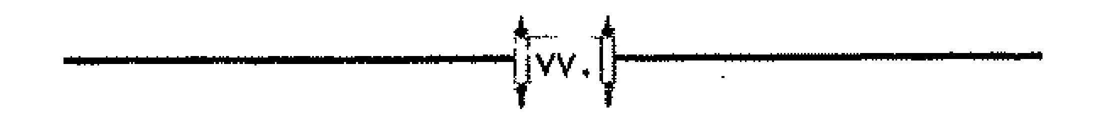
211 - Salvia, J. & Yasseldy Key, T. : Assessment in Special Remedial Education. Boston: Houghton Mifflin Co. , 1978.

212 - Sandven, J. Projectometry. Oslo: Universitesforlaget, 1975.

213 - Sax, G. Principles of Educational and Psychological Measurement and Evaluation. Belmont, C.A. : Wadsworth, 1989.

214 - Schader , W . New Directions for Testing and Measurement: Measuring Achivement, Progress Over a Decade, San Francisco: Jossey-Bass, 1980.

215 - Schroder, H. & Scuedfeld, P. Personality Theory and Information Processing . N.Y. : The Ronald Press Co. , 1971.



216 - Semoenoff, B. Projective Techniques N.Y. : John Wiley, 1976.

217 - Shavelson, R. & Webb, N. Generalizability Theory: 1973-1980 . **British Journal of Mathematical and Statistical Psychology,** 1981.

218 - Shaw, M. & Wright, J. Scales for the Measurement of Attitudes. N.Y. : McGraw-Hill, 1967.

219 - Shontz , F . Research Methods in Personality. N.Y. : Appleton - Century - Crofts, 1965.

220 - Sleeman, D. & Brown, B. Intelligent Tutoring Systems. N.Y. : Academic Press, 1982.

221 - Smith, L. & Klass, C. Assessing the Whole child. In : H. Simons and J . Elliott , Rethinking Appraisal and Assessment. Milton Keynes: Open University Press, 1989.

222 - Smith , P . ; Kandall, L. and Hulin , C . The Measurement of Satisfaction in Work and Retirement: A Strategy for the Study of Attitudes. Ill. : Rand McNally, 1969.

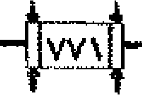
223 - Snow , R . "Attitude and Achievement" In New Directions for Testing . N.J. : ETS, 1980.

224 - Sonquist, J. & Dunkelbery, W. Survey and Opinion Research. N.J. : Prentice-Hall, Inc. , 1977.

225 - Srivastava , M . & Carter , E . An Introduction to Applied Multivariate statistics (3 rd ed.). N.Y. : North Holland, 1983.

226 - Stake, R. Advances in Program Evaluation . Greenwich: Jai Press Inc. , 1994.

227 - Stanley, J. & George , W. The Gifted and the Creative: A Fifty - Year Perspective . Baltimore : The Johns Hopkins University Press, 1977.



228 - Stephenson , W . The Study of Behavior : Q Technique and its Methodology. Ill. : University of Chicago Press, 1953.

229 - Sternberg, R. Intelligence, Information Processing, and Analogical Reasoning . N.J. : Lawrence Erlbaum Publishers, 1977.

230 - Sternberg, R. & Detterman, D. Human Intelligence: Perspectives on its Theory and Measurement. N.J. : ABLEX Publishing Corporation, 1979.

231 - Steven, R. "A Comparison of Item - and Person Fit Methods of Assessing Model - Data Fit in IRT " . **Applied Psychologi cal Measurement** 1990, 14 , 2 , 127 - 137.

232 - Stocking, M. & Swanson, L. "A Method for Severely Constrained Item Selection in Adaptive Testing " . **Applied Psychological Measurement,** 1993, 17 , 3 , 277-292.

233 - Stone , B. " Problems, Pitfalls, and benefits of Portfolies ". **Treacher Education Quarterly,** 1998 , 25 , 105-114.

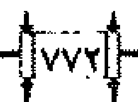
234 - Storms , B . ; Nuez, A. and Thomas, W. "Using Portfolios to Demonstrate Student Skills". **Thurst for Educational Leadership,** 1996, 25, 6-9.

235 - Sudman, S. & Bradburn , N. Asking Questions : A Practical Guide to Questionnaire Design. Calif. : Jossey-Bass, Publishers, 1982.

236 - Summers , G . Attitude Measurement. Ill. : Rand McNally & Company, 1971.

237 - Sundberg, N. Assessment of Persons. N.J. Prentice-Hall, Inc. , 1977.

238 - Tatsuoka, K. Rule Space: An Approach for Dealing With Misconceptions Based on Item Response Theory". **Journal of Educational Measurement,** 1983, 20, 4, 345-353.



239 - Tayler, C. Widening Horizons in Creativity. N.Y. : John Wiley, 1964.

240 - Tayler, I. & Getzels, J. Perspectives in Creativity. Ill . : Aldine Publishing Co. , 1975.

241 - Taylor, R. Assessment of Exceptional Students (3 rd ed.). Borton: Allyn and Bacon, 1993.

242 - TenBrink, T. Evaluation: A Practical Guide for Teachers. N.Y. : McGraw-Hill, 1974.

243 - Thorndike , R . Personnel Selection: Test and Measurement Techniques. N.Y. : John Wiley, 1961.

244 - Thorndike, R. Educational Measurement (2 nd ed.). Washington, D.C. : American Council on Education, 1971.

245 - Thorndike, R. Educational Measurement: Theory and Practice. In: D. Spearritt (Ed.). The Improvement of Measurement in Education and Psychology: Contributions of Latent Trait Theories. Hawthorn: ACER, 1982.

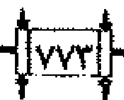
246 - Thorndike , R . & Hagen, E. Measurement and Evaluation in Psychology and Education . N.Y. : Wiley & Sons , Inc. 1969.

247 - Thurstone , L . The Measurement of Values. Ill. : University of Chicago Press, 1959.

248 - Thurstone , L. & Chave, E. The Measurement of Attitude . Ill. : The University of Chicago Press, 1929.

249 - Tiedman , H. Fundamentals of Psychological and Educational Measurement . Ill. : Charles Thomas Publisher, 1972.

250 - Tierney, R. ; Carter, M. and Desai, L. Portfolio Assessement in Reading - Writing Classroom. M.A. : Gordon, 1991.



251 - Toh, P. ; Swaminathan, G. "Application of Item Response Theory in Standard Setting". **Studies in Educational Evaluation,** 1989, 15, 285-293.

252 - Torrance, E. Guiding Creative Talent. N. J. : Prentice-Hall, 1962.

253 - Travers, R. Second Handbook of Research on Teaching . Ill. : Rand McNally & Company, 1973.

254 - Trow , C . & Haddan , E. Psychological Foundations of N.J.: Educational Technology Publications, Inc. , 1976.

255 - Tukey , J . Exploratory Data Analysis. Mass . : Addison-Wesley Publishing Company, 1977.

256 - Tyler, R. & Wolf, R. Crucial Issues in Testing. Calif. : McCutchan Corporation, 1974.

257 - UNESCO, International Conference on Education and Informatics. Paris: UNESCO, 1989.

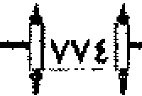
258 - United States Information Agency . "America 2000. The President Educational Strategy". U.S.A. , 1991.

259 - Urry, V . A Monte Carlo Investigation of Logistic Test Models . Unpublished Doctoral Dissertation, Princenton Universty, 1970.

260 - Vernon, P. Personslity Assessment. London: Methuen & Co. , Ltd. , 1965.

261 - Vernon , P. ; Adamson, G. and Vernon , D. The Psychology and Education of Gifted Children. London: Methuen & Co. Ltd. , 1977.

262 - Wallerg , H . & Haertel , G. The International Encyclopedia of Educational Evaluation. N.Y. : Pergaman Press, 1990.



263 - Walsh, W. & Tyler, L. Tests and Measurements, (3 rd ed.). N. J. : Prentice-Hall, 1979.

264 - Wechsler, D. WISC Manual . N.Y. : The Psychological Corporation, 1979.

265 - Weiner, E. & Stewart, B. Assessing Individuals: Psychological and Educational Tests and Measurements. Boston: Little, Brown and Company, 1984.

266 - Wenzel, L. ; Briggs, K.; and Puryear, B. "Portfolio: Authentic Assessment in the Age of the Curriculum Revolution". **Journal of Nursing Education,** 1998, 37, 5, 208-212.

267 - Wenzlaff, T. "Dispositions and Protfolio Development: is there a Connection?" **Education Journal, 1998, 118, 4, 564-572.**

268 - Wenzlaff, T. & Cummings, K. " The Portfolio as a Metapher for Teacher Reflection". **Contemporary Education,** 1996, 67, 109-112.

269 - Williams, P. Item Drift in Competency Based Programs. A Paper Presented at the Annual Meeting of the AERA , NewYork, 1982.

270 - Withers, G. & Batten , M. Defining Types of Assessment. In: B. Low and G. Withers (Eds.). Developments in School and Public Assessment. Hawthorn: ACER, 1990.

271 - Wolf, D. ; Bixby, J. ; Glenm, J. ; and Gardner, H. "To Use Their Minds Well: Investigating New Forms of Student Assessment". **Review of Research in Education, 1991, 16, 31-74.**

272 - Wood, R. Trait Measurement and Item Banks . In Gruijeter, D. & Kamp, L. Advances in Psychological and Educational Measurement . N. J. : Wiley, 1976.

273 - Wright, B. & Bell, S. Fair and Useful Testing With Item Banks. Chicago: MESA, Research Memo, 1980.

274 - Wright, B. & Douglas, G.. Better Procedures for Sample- Free Item Analysis. Research Memo No. 20 , University of Chicago, 1975.

275 - Wright, B. & Stone, M. Best Test Design : A Handbook for Rasch Measurement. Palo-Alto: Scientific Press, 1978.

276 - Zytowski, D. Contemporary Approaches to Interest Measurement. Minn. : University of Minnesota Press, 1973

A

Ability	قدرة
Ability Scale	ميزان القدرة
Ability Test	اختبار القدرة
Abscissa	المحور الأفقى / محور السينات
Absolute Zero	الصفر المطلق
Academic Aptitude	استعداد أكاديمى
Accomodation	مواءمة
Accountability	المسئولية التربوية
Achievement Test	اختبار تحصيل
Acquiscence	القبول/ الإذعان
Actuarial Prediction	التنبؤ الإحصائى
Adaptive Tests	اختبارات موائمة للفرد
Adjustment	تكيُّف
Aesthetic	متعلق بالذوق
Affect	وجدان
Affective Domain	المجال الوجدانى
Age Norms	معايير العمر
Age Scale	ميزان العمر (مقاييس الذكاء)
Alpha Coefficient	معامل ألفا (ثبات)
Alternate Forms	صيغ بديلة / متكافئة (اختبارات)
Ambiguity	غموض
Analogies Test	اختبار متشابهات
Analysis of Variance	تحليل التباين

Analysis of Covariance	تحليل التغاير
Anecdotal Record	سجل قصصى (لحياة الفرد)
Aptitude Test	اختبار استعداد
Arithmetic Mean	متوسط حسابى
Assembly Test	اختبار تجميع أشياء
Assessment	تقييم
Association	اقتران
Asymptote	خط تقاربى
Attenuation	انخفاض قيمة الارتباط بين متغيرين نتيجة عدم دقة عملية القياس
Attitude	اتجاه
Attitude Scale	مقياس (ميزان) اتجاه
Authentic Assessment	تقييم واقعى
Automated Scoring	تصحيح آلى

B

Basal Age	العمر القاعدى (مقاييس الذكاء)
Basic Skills	مهارات أساسية
Battery of Tests	بطارية اختبارات
Bayes Statistics	إحصاء بيز
Behavioral Domain	نطاق سلوكى
Behavioral Objective	هدف سلوكى
Bias	تحيُّز
Bimodal Distribution	توزيع ثنائى المنوال
Binomial Distribution	توزيع ذى الحدين
Bipolar	ثنائى القطب

Biserial Correlation	معامل الارتباط الثنائى المتسلسل
Bivariate Distribution	توزيع ثنائى المتغير
Branching Model	نموذج تفريعى (اختبارات)

C

Causality	علّية / سببية
Checklist	قائمة مراجعة
Chi- Square test	اختبار مربع كاى (إحصاء)
Chronological Age	العمر الزمنى
Classification	تصنيف
Clerical Aptitude	الاستعداد الكتابى
Clinical Prediction	التنبؤ الكلينيكى
Cognition	معرفة
Cognitive Processes	عمليات معرفية
Cognitive Psychology	علم النفس المعرفى
Cognitive Styles	أساليب معرفية
Common Scale	ميزان مشترك (اختبارات)
Competency - Based Evaluation	تقويم يستند إلى كفايات
Competency Test	اختبار كفاية
Completion Item	مفردة تكملة
Composite Score	درجة مركبة
Computer Simulation	المحاكاة باستخدام الحاسوب
Computerized Spreadsheat	مصفوفة بيانات الكترونية
Concurrent Validity	الصدق التلازمى
Conditional Probability	الاحتمال المشروط
Conjoint Measurement	القياس المتّحد

Construct Validity	صدق التكوين الفرضى
Content Validity	صدق المحتوى
Convergent Thinking	التفكير التقاربى (النمطى)
Correction for Guessing	التصحيح من أثر التخمين
Correlation	معامل الارتباط
Cost / Benefit Ratio	نسبة الكلفة إلى العائد
Counseling	إرشاد
Covariance	تغاير
Criterion Groups	مجموعات محكّة
Criterion - Referenced Test	اختبار مرجعي المحك
Criterion - Related Validity	الصدق المتعلق بمحك
Critical Incidents	أحداث حرجة
Critivity Test	اختبار ابتكار
Cross - Validation	الصدق المستعرض
Crystalized Intelligence	الذكاء المتبلور
Culture - Fair Test	اختبار غير متحيز لثقافة
Culture - Free Test	اختبار متحرر من الثقافة
Cumulative Frequency	تكرار متجمع
Cumulative Model	نموذج تراكمى (نموذج جتمان)
Cut-off Score	درجة القطع



D

Decile	إعشارى
Decision Studies	دراسات القرارات (اختبارات)
Decision Theory	نظرية القرارات
Dependent Variable	متغير تابع

Derived Score	درجة مشتقة (مُحولة)
Determination Coefficient	معامل التحديد (إحصاء)
Deterministic Models	نماذج حتمية
Deviation IQ	نسبة الذكاء الانحرافية
Diagnostic Test	اختبار تشخيصى
Dichotomous	ثنائى
Difference Score	درجة الفرق
Differential Psychology	علم النفس الفارق
Differentiation	تمايز
Discrete Variable	متغير منفصل
Discrimination Index	معامل التمييز (اختبارات)
Dispersion	تشتت
Disposition	تهيؤ / استعداد
Distractors	مُشتّتات (اختبارات)
Divergent Thinking	تفكير تباعدى (منطلق)
Domain - Referenced Test	اختبار مرجعى النطاق
Domain - Selected Validity	صدق انتقاء النطاق السلوكى

E

Educational Profile	صفحة تربوية/ بروفيل تربوى
Elaboration	تحسين / إحكام
Empirical Validity	الصدق الإمبيريقى
Equivalent Form	صيغة مكافئة (اختبارات)
Errors of Estimate	أخطاء التقدير
Errors of Measurement	أخطاء القياس
Error Variance	تباين الخطأ

Essay Test	اختبار مقال
Evaluation	تقويم
Expectancy Table	جدول توقُّع
Experimental Design	تصميم تجريبى
Expressed Interests	ميول مُعبَّر عنها
External Criterion	محك خارجى
Exterpolation	استكمال (إحصاء)
Extravert	منبسط / اجتماعى

F

Face Validity	الصدق الظاهرى
Facet	وجْه / بُعد (اختبارات)
Factors	عوامل
Factor Analysis	التحليل العاملى
Factor Loadings	تشبُّعات عاملية
Fairness	عدالة
False Negatives	أخطاء سالبة
False Positives	أخطاء موجبة
Feedback	تغذية مرتدة
Field- Dependent	معتمد على المجال الإدراكى
Figural	متعلق بالأشكال / شكلى
Flexilevel Test	اختبار ذو مستوى مرن
Fluency	طلاقة
Fluid Intelligence	الذكاء المائع
Forced- Choice Items	مفردات جبرية الاختيار
Formative Evaluation	تقويم بنائى

Frequency Distribution	توزيع تكرارى
Frequency Polygon	مضلع تكرارى
Function	دالة/ وظيفة
Functional Validity	الصدق الوظيفى

G

General Factor	عامل عام
Generalizability Coefficient	معامل إمكانية التعميم (اختبارات)
Generalizability Studies	دراسات إمكانية التعميم (اختبارات)
Generalizability Theory	نظرية إمكانية التعميم (اختبارات)
Genetic	وراثى
Goal-Free Evaluation	تقويم لا يستند إلى الأهداف
Goodness of Fit	جودة المطابقة (إحصاء)
Grade Norms	معايير الفرق الدراسية
Grade Point Average (GPA)	المعدَّل العام
Graded Reponse Model	نموذج الاستجابة المتدرِّجة
Graphic Rating Scale	ميزان تقدير بيانى
Group Factor	عامل طائفى
Group Test	اختبار جماعى
Guessing	تخمين (اختبارات)
Guidance	توجيه

H

Halo Effect	تأثير الهالة
Hawthorn Effect	تأثير هاوثورن (تصميمات تجريبية)
Heuristic Models	نماذج كشفية

Histogram	مدرج تكرارى
Homogeneity	تجانس
Horizontal Equating	التكافؤ الأفقى (اختبارات)
Hypothesis	فرض

I

Imagery	تصور (عقلى)
Implicit Objectives	أهداف ضمنية
In- Basket Items	مفـردات مواقـف تُحاكـى المهـام الإدارية
Incremental Validity	الصدق المتزايد
Independent Variable	متغير مستقل
Individual Test	اختبار فردى
Information Function	دالة المعلومات (اختبارات)
Information Processing	تجهيز المعلومات ومعالجتها
Intelligence	الذكاء
Intelligence Quotient	نسبة الذكاء
Intelligence Test	اختبار ذكاء
Interest Inventory	استبيان ميول
Interlocking Items	مفردات متداخلة (اختبارات)
Internal Consistency	اتساق داخلى (اختبارات)
Interpolation	استكمال من الداخل
Interval Scale	ميزان فترى
Interview	مقابلة
Introvert	منطوٍ
Invariance	عدم التغير
Inventory	استبيان

Inventorized Interests	ميول مقاسة بالاستبيان
Ipsative Measurement	قياس مرجعى الفرد
Item	مفردة (اختبارات)
Item Analysis	تحليل المفردات
Item Bank	بنك أسئلة
Item Bias	تحيز المفردة
Items Calibration	تعيير المفردات
Item Characteristic	المنحنى المميز للمفردة
Item Difficulty	صعوبة المفردة
Item Discrimination	تمييز المفردة
Items File	ملف المفردات
Item Forms	صيغ المفردات
Item Generation Rules	قواعد تكوين المفردات (اختبارات)
Item Links	وصلات المفردات
Item Logit	الترجيح اللوغاريتمى للمفردة (اختبارات)
Item Networks	شبكات المفردات
Items Pool	مجموعة من المفردات
Item Response Theory (IRT)	نظرية الاستجابة للمفردة الاختبارية

J

Job Analysis	تحليل العمل

K

Kappa Coefficient	معامل كابا (إحصاء)

Kuder - Richardson Formula	صيغة كيودر- ريتشاردسون (اختبارات)

L

Latent Class Model	نموذج الفئات الكامنة (نموذج لازرسفلد)
Latent Trait Theory	نظرية السمات الكامنة
Latent Variable	متغير كامن (سمة أو قدرة)
Liniency Error	خطأ متعلق بالتساهل (فى التقديرات)
Likelihood Function	دالة الأرجحية (إحصاء)
Likert Method	أسلوب ليكرت فى قياس الاتجاهات
Linear Regression	انحدار خطى
Local Independence	الاستقلال المركزى (إحصاء)
Local Norms	معايير محلية
Logistic Models	نماذج ترجيح لوغاريتمى
Logit	وحدة ترجيح لوغاريتمى (اختبارات)

M

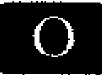
Machine - Scored Test	اختبار يصحَّح آليًا
Main Effect	تأثير أساسي (إحصاء)
Mastery Test	اختبار تمكُّن
Matching Item	مفردة مزاوجة (اختبارات)

Matrix Sampling	معاينة المصفوفة (اختبارات)
Maximum Likelihood	الأرجحية القصوى (إحصاء)
Mean Squares	مجموع المربعات (إحصاء)
Measurement	قياس
Median	الوسيط
Mental Age	العمر العقلى
Meta- Evaluation	التقويم الفوقى
Mode	المنوال
Moderater Variable	متغير وسيط
Monotonic Relationship	علاقة اطرادية
Multidimensional Models	نماذج متعددة الأبعاد
Multidimensional Scaling	موازين متعددة الأبعاد
Multifacet	متعدد الأوجه
Multiple- Choice item	مفردة اختيار من متعدد (اختبارات)
Multiple Correlation Coefficient	معامل الارتباط المتعدد (إحصاء)
Multiple - Factor Theory	نظرية العوامل المتعددة
Multiple - Level Test	اختبار متعدد المستويات
Multitrait - Multimethod Matrix	مصفوفة السمات المتعددة والطرق المتعددة

N

National Norms	معايير قومية
Needs Assessment	تقدير الاحتياجات
Nested Design	تصميم تجريبى متشابك (متداخل)
Nominal - Response Model	نموذج الاستجابة الاسمية (التصنيفية)

Nominal Scale	ميزان قياس اسمى (تصنيفى)
Nonlinear Equating	التكافؤ غير الخطى (اختبارات)
Nonparametric Statistics	إحصاء لابارامترى (معلمى)
Nonverbal Test	اختبار غير لفظى
Norm Group	جماعة مرجعية (معيارية)
Normal Distribution	التوزيع الاعتدالى
Normalized Scores	درجات مُعدَّلة اعتداليًا
Norm - Referenced Test	اختبار مرجعى الجماعة أو المعيار
Norms	معايير



Objectives	أهداف
Objective Test	اختبار موضوعى
Observation Method	أسلوب الملاحظة
Obtained Scores	درجات ملاحظة
Odd - Even Reliability	التجزئة النصفية (ثبات)
Ogive	منحنى تكرارى متجمع
Omnibus Test	اختبار شامل
Oral Test	اختبار شفوى
Ordinal Scale	ميزان قياس رتبى
Ordinate	المحور الرأسى/ محور الصادات
Originality	الأصالة



Paired Comparison Method	طريقة المزاوجة الثنائية
Paper - and Pencil Test	اختبار ورقة وقلم



Parallel Forms	صيغ متوازية (اختبارات)
Paramater	بارامتر (معلم)
Peer Ratings	تقديرات الأقران
Percentiles	مئينيات
Percentile Band	مدى مئينى
Percentile Norms	معايير مئينية
Percentile Rank	رتب مئينية
Perception	إدراك
Performance Contract	عقد أداء
Performance Standard	مستوى أداء
Performance Test	اختبار أداء
Personality	الشخصية
Personality Inventory	استبيان الشخصية
Personality Test	اختبار الشخصية
Phi Coefficient	معامل فاى (إحصاء)
Placement Test	اختبار تسكين
Point - Biserial Corelation	معامل الارتباط الثنائى المتسلسل الحقيقى (إحصاء)
Point Scale	ميزان قياس يعتمد على النقط
Population	مجتمع
Portfolio	صفحة وثائقية
Portfolio Assessment	التقييم التوثيقى الشامل
Power Test	اختبار قوة
Practice Effect	تأثير التدرب أو الممارسة
Prediction	تنبؤ
Predictive Validity	الصدق التنبؤى
Prior Information	معلومات قبلية

Probabilistic Models	نماذج احتمالية
Profile (Psychograph)	صفحة نفسية (بروفيل نفسى)
Program Evaluation	تقويم البرنامج
Prognosis	تنبؤ بنتائج
Progressive Matrices	مصفوفات متتابعة
Projective Techniques	أساليب إسقاطية
Projective Test	اختبار إسقاطى
Projectometry	القياس الإسقاطى
Proximity Data	بيانات متماثلة (مقترنة)
Psychomotor	نفسحركى

Q

Q - Sort	التصنيف الترتيبى (Q)
Qualitative	كيفى / نوعى
Quantitative	كمّى
Quartile	إرباعى
Questionnaire	استبيان

R

Random Effects Model	نموذج التأثيرات العشوائية (إحصاء)
Random Sample	عينة عشوائية
Range	مدى
Rapport	صلة وئام / أو ألفة
Rasch Model	نموذج راش (النموذج أحادى البارامتر)

Rating Scale	ميزان تقدير
Ratio Scale	ميزان قياس نسبى
Raw Score	درجة خام
Reactive	تفاعلى (راد لتأثير الفعل)
Readiness Test	اختبار تهيؤ
Recall Item	مفردة تتطلب استدعاء (اختبارات)
Recognition Item	مفردة تتطلب تعرُّف (اختبارات)
Rectangular Distribution	توزيع مستطيل (إحصاء)
Regression Equation	معادلة انحدار (إحصاء)
Regression Line	خط انحدار (إحصاء)
Reliability	الثبات
Reliability Coefficient	معامل الثبات
Representative Sample	عينة مُمثلة
Residuals	بواقى
Residual Error	بواقى الخطأ (إحصاء)
Response Bias	تحيز الاستجابة
Respone Sets	اتجاهات عقلية
Response Styles	أساليب الاستجابات (اختبارات)

S

Sample	عينة
Sampling Theory	نظرية المعاينات
Scaling	عملية تحويل الدرجات إلى موازين
Scatter Diagram	شكل انتشارى (بين متغيرين)
Schema	مُخطط عقلى
Scholastic Aptitude Test	اختبار استعداد دراسى

Screening Test	اختبار فحص
Selection	انتقاء
Selection Ratio	نسبة الانتقاء
Self - Concept	مفهوم الذات
Self - Report Inventory	استبيان تقرير ذاتى
Semantic Differential	التمايز اللغوى
Semi - Interquartile Range	نصف المدى الرُبيعى (إحصاء)
Sentence Completion Test	اختبار تكملة الجمل
Sequential Test	اختبارات تتابعية
Situational Test	اختبار مواقف
Skewness	التواء
Social Desirability	مراعاة القبول الاجتماعى
Sociogram	شكل بيانى للعلاقات الاجتماعية
Sociometric Matrix	مصفوفة علاقات اجتماعية
Sociometry	القياس الاجتماعى
Software	برمجيات حاسوب
Source Traits	سمات مصدر
Spatial Ability	قدرة مكانية
Spearman-Brown Prophecy Formula	صيغة سبيرمان وبراون (ثبات)
Specific Determineres	محدِّد نوعى (لمفردة اختيار من متعدد)
Speed Test	اختبار سرعة
Spiral Omnibus Test	اختبار حلزونى شامل
Split-Half Reliability	الثبات بطريقة التجزئة النصفية (اختبارات)
Stability Coefficient	معامل الاستقرار (اختبارات)

Standard Deviation	الانحراف المعيارى (إحصاء)
Standard Error of Measurement	الخطأ المعيارى للقياس
Standard Scores	درجات معيارية
Standardization	تقنين
Standardization Group	عينة التقنين
Standardized Test	اختبار مُقنن
Stanine	التساعى المعيارى
Static Models	النماذج الاستاتيكية (السكونية)
Statistic	إحصاء
Stem (of an Item)	العبارة الأساسية (مفردة اختيار من متعدد)
Stem - and - Leaf	شكل الأغصان والأوراق (إحصاء)
Stratified Sample	عينة طبقية
Structure of Intellect	بنية العقل
Subtset	اختبار فرعى
Success Ratio	نسبة النجاح
Summated Ratings	تقديرات تجميعية (ميزان ليكرت)
Summative Evaluation	تقويم ختامى
Suply Item	مفردة تتطلب إجابة حرة (اختبارات)
Surface Traits	سمات سطحية
Survey Test	اختبار مسحى

T

T Scores	الدرجات التائية
Tab Test	اختبار مهام متتابعة لتشخيص مشكلة واقعية وحلها

Table of Specifications	جدول مواصفات
Tailored Tests	الاختبارات المفصَّلة
Task Analysis	تحليل المهام
Taxonomy	تصنيف
Teacher - Made Test	اختبار يكتبه المعلم (اختبار صفِّى)
Test	اختبار
Test Administration	تطبيق (إجراء) الاختبار
Test Anxiety	قلق الاختبار
Tests Networks	شبكات الاختبارات
Test - Retest Reliability	الثبات بطريقة إعادة تطبيق الاختبار
Test Security	سرِّية الاختبار
Tetrachoric Correlation	معامل الارتباط الرباعى (إحصاء)
Threshold Score	درجة فارقة (اختبارات)
Time Sampling	معاينة زمنية
Trait	سمة
True - False Item	مفردة صواب أو خطأ (اختبارات)
True Score	درجة حقيقية
True Variance	تباين حقيقى



Unidimensional	أحادى البُعد
Unconditional Probability	احتمال غير مشروط (إحصاء)
Universe	نطاق شامل
Universe Score	درجة شاملة (اختبارات)

V

Validity	الصدق
Variability	تغيُّر
Variance	تباين
Venn Diagram	شكل فن (لتمثيل المجموعات)
Verbal Test	اختبار لفظى
Vertical Equating	التكافؤ الرأسى (اختبارات)
Vocational Interest Test	اختبار ميول مهنية

W

WAIS	مقياس ويكسلر لذكاء الراشدين
WISC	مقياس ويكسلر لذكاء الأطفال
WITS،	وحدة ترجيح لوغاريتمى مُعدلَّة (اختبارات)
Word Association Test	اختبار تداعى الكلمات
WPIS	مقياس ويكسلر لذكاء أطفال ما قبل المدرسة

Z

Z Score	درجة معيارية

محتويات الكتاب

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

رقم الإيداع	٩٩/٢١٠٣
I. S. B. N الترقيم الدولي	977- 10 -1210-0

مراجع أخرى للمؤلف

(١) اتجاهات معاصرة فى تصميم وبناء نظم بنوك الأسئلة.

(٢) الاختبارات التشخيصية مرجعية المحك فى المجالات التربوية والنفسية والتدريبية **«دار الفكر العربى»**.

(٣) اختبار تشخيصى مرجعى المحك لقياس مهارات المعلمين فى بناء الاختبارات المدرسية.

(٤) الأساليب الإحصائية الاستدلالية البارامترية واللابارامترية فى تحليل بيانات البحوث النفسية والتربوية **«دار الفكر العربى»**.

(٥) تحليل البيانات فى البحوث النفسية والتربوية **«دار الفكر العربى»**.

(٦) تحليل المحتوى وتقويمه فى مرحلة التعليم الأساسى.

(٧) تطورات معاصرة فى القياس النفسى التربوى.

(٨) تقويم البحوث التربوية فى دول المشرق العربى.

(٩) دليل المعلم فى تصميم وبناء الاختبارات التحصيلية الصفية المعاصرة.

(١٠) دليل المعلم فى تقويم الطلبة فى الدراسات الاجتماعية **«دار الفكر العربى»**.

(١١) مفاهيم وممارسات التعليم الأساسى فى بعض الدول غير العربية.

(١٢) نظم وأساليب التقويم والامتحانات فى الدول العربية: الواقع والتوجهات المستقبلية.

هذا الكتاب

يعد مرجعا أساسيا للدارسين والباحثين والممارسين في مختلف مجالات القياس والتقويم التربوي والنفسي، حيث يضم في مجلد واحد الأسس النظرية المنهجية والسيكومترية، والتطبيقات المتعددة، والتطورات والتوجهات المعاصرة المتعلقة بهذه المجالات.

ويشتمل الكتاب على أربعة أبواب رئيسة. يتناول **الباب الأول** الذي يشتمل على سبعة فصول المفاهيم والمبادئ الأساسية للقياس والتقويم، والخصائص الإحصائية والسيكومترية للاختبارات والمقاييس، والأساليب المتعددة لتقدير صدقها وثبات درجاتها، وتحديد معاييرها، وتحليل مفرداتها، ومشكلات القياس التربوي والنفسي وشروطه العلمية وأخلاقياته.

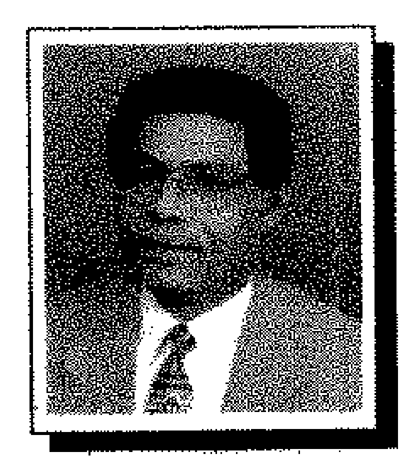
ويتناول **الباب الثاني** الذي يشتمل على ثلاثة فصول منهجيات وأساليب قياس وتقويم الجوانب المعرفية المتمثلة في التحصيل الدراسي والكفايات، والذكاء الإنساني، والاستعدادات الخاصة بأنواعها المختلفة.

ويتناول **الباب الثالث** الذي يشتمل على ثلاثة فصول أيضا، منهجيات وأساليب قياس وتقويم الجوانب غير المعرفية المتمثلة في الميول، والاتجاهات، والشخصية الإنسانية.

ويشتمل كل من هذين البابين على أمثلة متنوعة من الاختبارات والمقاييس الشائعة الاستخدام في قياس الجوانب المعرفية، والنفسية الحركية، والوجدانية، والشخصية.

أما **الباب الرابع** فيشتمل على فصلين يتناول أحدهما أهم التطورات النظرية المعاصرة في القياس والتقويم التربوي والنفسي، ويتناول الآخر بعض التوجهات التطبيقية المعاصرة التي تستند إلى هذه التطورات النظرية.

ويعالج الكتاب الجوانب النظرية والتطبيقية معالجة متكاملة، حيث تستند التطبيقات المختلفة إلى الأطر النظرية المتعددة، ونتائج الدراسات والبحوث المتعلقة بها. كما يوضح المفاهيم والأساليب المختلفة بالعديد من الأشكال التخطيطية لكي يتيسر على القارئ استيعابها واستخدامها في مجالات اهتمامه. وينتهي الكتاب بمعجم مختصر للمصطلحات الأساسية في القياس والتقويم التربوي والنفسي.



د. صلاح الدين محمود علام

* بكالوريوس في الرياضيات والتربية من جامعة عين شمس عام ١٩٦٠م.
* ماجستير في التربية (علم نفس تعليمي) من جامعة عين شمس عام ١٩٧١م.
* دكتوراه الفلسفة في التربية تخصص قياس وتقويم وإحصاء نفسي وتربوي من جامعة ميتشجان الأمريكية عام ١٩٨٠م.
* أستاذ القياس والتقويم والإحصاء بكلية التربية جامعة الأزهر.
* مستشار القياس والتقويم في عدد من المؤسسات والمنظمات التربوية العربية.
* صمم وأشرف على تنفيذ العديد من المشروعات التطويرية لنظم وأساليب التقويم والامتحانات في الدول العربية.
* أجرى العديد من البحوث والدراسات في المجالات الإحصائية والسيكومترية المعاصرة، وله عدة مؤلفات أخرى في هذه المجالات.
* قام بالإشراف على عدد من رسائل الماجستير والدكتوراه في مجالات القياس مرجعي المحك، والنماذج السيكومترية المعاصرة.
* عضو الجمعية الأمريكية للبحث التربوي، والمجلس القومي الأمريكي للقياس التربوي، وعضو مؤسس للمجلس العربي للموهوبين والمتفوقين.